SCIENCE NEWS

EDITED BY A. W. HASLETT

28

PENGUIN BOOKS

MELBOURNE · LONDON · BALTIMORE

BIOLOGICAL CONTROL OF THE RABBIT IN AUSTRALIA

FRANK FENNER AND M. F. DAY

THERE are several examples of the successful control of plant and insect pests by means of insect or microbial parasites, the most spectacular being the control of prickly pear in Australia by an imported Argentine moth *Cactoblastis cactorum*.* But biological control of vertebrate pests has seldom been attempted, and has so far been unsuccessful. The forerunner in this as in so many other fields of microbiology, it was Pasteur who first attempted to destroy animals with a microbial parasite. He spread a culture of chicken-cholera bacteria on cut lucerne around the burrows of rabbits on the estate of Madame Pommery, the owner of the famous champagne firm, and within a fortnight all the rabbits were dead. Attracted by the £25,000 reward offered by the Government of New South Wales in 1887 for the demonstration of an effective method of exterminating rabbits, Pasteur sent his technical assistant Loir to Australia to attempt distribution of chicken-cholera cultures. He hoped, rather naïvely, to repeat his success in France by the same method of distribution of the bacteria. However, permission to bring the culture into the country was refused, as it was highly pathogenic for the domestic fowl as well as other animals. In 1907, trials were made in New South Wales with a related infection of rabbits recommended by Dr. Danysz, of the Pasteur Institute, but it failed to establish itself in preliminary trials. Later it was shown that infection with this bacterium was enzootic in Australian rabbits.

During the last two years, however, after many years of preliminary experimentation and several unsuccessful field trials,

*See, for example, Wilson, F. *New Biology* **8**, 51 (1950).

the virus of infectious myxomatosis of rabbits has been artificially introduced at a number of places in New South Wales and Victoria, and has spread over the greater part of south-eastern Australia, causing an enormous mortality amongst the wild rabbits which are such a serious pest there. This is an event of considerable importance, not only for its promise in the struggle to control the rabbit pest in Australia, but in its broader significance as an example of the successful establishment and spread of a highly lethal microbial infection in mammals. The rapidity and extent of the initial spread of the disease in the summer of 1950–51 was so remarkable – it extended over an area of roughly 1,000 × 700 miles within three months of its first recognition away from the sites where it was released – that it is worth while trying to assess the factors responsible for its dissemination. We attempt this in the following pages.

Before describing the myxomatosis outbreak itself, we must examine two of the principal characters, the rabbit and the virus. Later we shall see that the activity of a third character must be studied if we are to understand the spread of the disease.

THE ECOLOGY OF THE RABBIT IN AUSTRALIA

Much of the early history of rabbits in Australia has been summarized by D. G. Stead in his booklet *The Rabbit in Australia*, published in 1935. The first five rabbits to be introduced arrived in 1788 with the First Fleet under Governor Phillip. These, however, do not seem to have prospered. Although rabbits were listed in 1824 among the domestic animals of van Diemen's Land (now Tasmania), and were recorded as occurring in thousands within a few years, they do not seem to have spread rapidly on the mainland until more than thirty years later. In 1859 a farmer near Geelong, Victoria, received 24 rabbits. Six years later he killed 20,000 on his property and estimated that one-third of the population remained. This seems to have been where the rabbits really got their start. In the late 1870's they crossed the Murray River into New South Wales, within ten years they had reached Queensland, and by 1895 they had crossed from South Australia into Western Australia. It is

interesting to reflect that myxomatosis covered in about three months an area of eastern Australia roughly the same as that occupied by rabbits in fifteen to twenty years.

Within forty years rabbits had spread from negligible numbers in Victoria to cover an area roughly equivalent to half of the continent of Europe. This rapid increase depended on a number of factors. These include the rapid reproductive rate (Stead records that rabbits reach sexual maturity at four months, produce an average of six litters per year with an average of five or six young per litter), the relative absence of natural enemies and of native animals occupying the same ecological niche, and the abundant food supply. The rabbit has never successfully colonized the northern parts of the continent, and the northern limit of its distribution, which approximates the 23rd parallel of latitude, varies with seasonal conditions.

The numbers of rabbits also fluctuate with the seasons, as can be seen from figures for exports of carcasses and skins. The slaughter, mostly by trapping, of some hundred million rabbits every year has, because of the animal's fecundity, little appreciable effect on their numbers in lush seasons, and many additional methods of control have been attempted. These can only be listed here, and include poisoning, particularly with strychnine and phosphorus, fumigating with hydrogen cyanide, exhaust fumes, etc., driving, digging and ripping of warrens, the use of ferrets and dogs, and shooting. Many other methods have been suggested, but none have been of much value and, in spite of all that has been done, the rabbit problem in Australia has so far defied all efforts to combat it.

THE VIRUS OF INFECTIOUS MYXOMATOSIS

The wild rabbits of Australia all belong to the species *Oryctolagus cuniculus*, as do the domestic rabbits of Europe and the rabbits used in biological laboratories all over the world. This is the only species of rabbit occurring in Europe and Australia, but there are several species in North and South America, all belonging to other genera. When bacteriologists set up laboratories in America they brought the European rabbit across as a

laboratory animal. First in South America, and much later in California, a strange disease broke out at irregular intervals in these laboratory rabbits, and in other commercial rabbitries which also bred European rabbits. Every animal which contracted the disease died.

In 1898 an Italian bacteriologist, G. Sanarelli, who was working in Uruguay, recovered a filterable virus from the diseased animals of his laboratory stock and named the disease infectious myxomatosis. It was, incidentally, the second filterable virus of animals to be characterized, that of foot-and-mouth disease having been described by Loeffler and Frosch a few months earlier. For the next fifty years outbreaks of this disease, universally lethal to affected rabbits, broke out in various parts of South America, in Mexico, and in Southern California, but its origin remained obscure. It was not until 1943 that Professor Aragão, in Rio de Janeiro, showed that a large proportion of the common wild rabbits of Brazil were immune to infection with the virus, not, as earlier workers had thought, because of an innate resistance, but because they had recovered from prior infection. In susceptible rabbits of the genus *Sylvilagus* inoculation of the virus resulted in the development of a localized swelling of the skin, which slowly regressed. The disease rarely became generalized and was never fatal. However, mosquitoes which bit through the infected skin of these rabbits were able to convey both the benign disease to other *Sylvilagus*, and the lethal disease to laboratory rabbits. Thus the natural disease is a benign infection of the native rabbits of certain parts of America, and is transmitted from one rabbit to another by insect vectors.

What does the virus itself look like? Photographs taken with the electron microscope by J. Farrant (Inset 21) show a 'brick-shaped' particle which is practically identical in size and shape with the virus particles of the pox diseases of men, animals and birds; see inset 22, which shows the vaccinia virus for comparison. Indeed, some investigators think that myxoma virus must be grouped with the pox viruses in any scheme of classification, although there are no immunological relationships between any of the pox viruses and myxoma virus. However, there is a more

closely related disease of cottontail rabbits (*Sylvilagus*) in eastern North America which closely resembles myxomatosis of the South American wild rabbit in that it merely causes single, large, tumour-like swellings to develop in infected animals. Dr Richard Shope, working at the Rockefeller Institute at Princeton, New Jersey, isolated a virus from these tumours, and when laboratory rabbits were inoculated with this virus they developed small lumps only. In the domestic rabbit, no two diseases could be less alike, superficially, than this disease and myxomatosis. Shope's fibroma virus caused no general symptoms, but merely the development of a small lump at the inoculation site. Sanarelli's myxoma virus caused a generalized disease which was fatal in an extraordinarily high proportion of infected rabbits, and which was characterized by great swelling of the eyes and head (which have earned the disease the name of 'bighead' in California) and the development, in animals which did not die too quickly, of lumps all over the body. Yet Shope found that previous inoculation of domestic rabbits with fibroma virus protected them against the lethal effects of myxomatosis, and other investigators have repeatedly confirmed this.

On the continent of America, therefore, there appear to be two closely related virus diseases of native rabbits, caused by two viruses which may be the American representatives of the pox viruses of the Old World. One causes a trivial disease in the foreign European rabbit, the other is almost invariably lethal for it.

Before we proceed to describe early trials of rabbit destruction with myxoma virus, it is important to emphasize the unique features of the disease, namely the almost invariably fatal outcome of infection in domestic rabbits and the failure to cause disease in any other species of animal or bird. From one laboratory in the United States it is recorded that there was not one recovery amongst 2,000 rabbits infected, and other workers in England and Australia record figures such as one survivor in over 200 rabbits infected by contact, and two recoveries out of 929 exposed to contact infection. Very recent investigations suggest that under natural conditions the case-mortality rate in

Australia, in populations which had never before been exposed to the disease, was of the same order, viz. 99.5 per cent to 99.8 per cent. The very high mortality rate of myxomatosis in the European rabbit is better appreciated if we consider a few human infectious diseases which we consider highly fatal. The most lethal infectious disease of man is undoubtedly plague, in which the mortality rate in the pneumonic form approaches 100 per cent, and in the more common bubonic form is from 60 per cent to 90 per cent. Here we have a situation which is comparable to myxomatosis of the European rabbit, in that man is only an incidental host for the bacterium, which survives in nature in a large variety of rodents, in which its lethality is much lower. Two specifically human diseases usually associated with high mortality rates are pandemic influenza and cholera. In the last great influenza pandemic the attack rate in closed communities in which all individuals were exposed to the disease rarely exceeded one half, and the case-mortality rate even in the most susceptible age-group of such communities was usually only from 5 to 10 per cent. The case-mortality rate in cholera in India rarely exceeded 70 per cent, even in the older and more susceptible age-groups, and in the great Hamburg epidemic of 1892 it was 46 per cent.

There are, of course, good biological reasons for the rarity of this extreme manifestation of a parasite-host relationship. Such a 'balance' is incompatible with the evolutionary survival of either host or virus. We will examine this point in greater detail when we discuss the possible future of the use of myxomatosis in rabbit control.

The second important characteristic of the virus is the narrow range of susceptible hosts, which are limited to the European rabbit, certain native American rabbits of the genus *Sylvilagus*, and developing embryos of chickens and ducks.

THE USE OF MYXOMATOSIS FOR RABBIT DESTRUCTION: PRELIMINARY EXPERIMENT AND TRIALS

The first suggestion that myxomatosis might be used for the destruction of rabbits came from Professor Aragão of Brazil,

in 1927, and it was followed up by a few experiments in Australia. These experiments showed that, although the disease could be transferred by contact, burrows recently vacated by diseased rabbits were not infectious. As the latter requirement was judged to be essential if the disease was to be of any practical use, no further trials were made in Australia at that time.

A few years later Sir Charles Martin, who had then retired from the Directorship of the Lister Institute, undertook some important investigations at Cambridge, England, with the intention of determining whether myxomatosis might be useful for the control of rabbits. The essential requirements of a disease organism which is to be used for the control of vermin are as follows. The micro-organism must produce a disease which will spread through the population of the animal to be destroyed, and it must be of such high virulence for that animal that few survive the infection to build up a resistant race. Further, it must maintain this high degree of virulence during natural passage, or the epizootic will 'peter out'. If any animals do recover, their progeny must not inherit sufficient resistance to allow them also to survive infection. Moreover, the degree of host specificity of the agent must be high, for it must be harmless for all species of animal and bird, both wild and domesticated, other than the pest it is desired to destroy. After a thorough examination of the scientific literature on the subject, and experimentation extending over twenty months, and including four experiments with colonies of rabbits living under almost natural conditions, Martin concluded that the virus of myxomatosis satisfied these essential requirements.

The next step was field experiments on a larger scale in Australia, and these were commenced by the Council of Scientific and Industrial Research of the Commonwealth of Australia in 1936. First, more extensive experiments were carried out to ensure that the virus was harmless not only to domesticated birds and animals. but to the native fauna of

trials were made on an enclosure of 90 acres on a small island, one trial extending through the summer and the other through the winter. Although a number of uninoculated rabbits died of myxomatosis on each occasion, the disease failed to exterminate the rabbits in the 90-acre enclosure.

In the meantime laboratory experiments carried out by Dr L. B. Bull had shown that the stickfast flea, *Echidnophaga*, could carry the disease from one rabbit to another. A field experiment was therefore tried in a locality where these parasites infested the rabbits. In an enclosure of 90 acres, during the months of January and February, the disease spread to all warrens in the enclosure and exterminated a population of about 500 rabbits, except for 17 which failed to contract the infection. Field trials were then set up in semi-arid bush country in the north-east portion of South Australia, under completely natural conditions, which included the presence of foxes and other predatory animals. Favourable results were obtained when foxes were not present in appreciable numbers, and especially when the rabbit population was infested with the native stickfast flea, but for the most part the disease quickly died out. Further work was then suspended until the end of the 1939–45 war.

It was obvious that field trials were necessary in well-watered country, for both Bull in Australia and Aragão in South America had shown that mosquitoes could convey the disease. These trials were begun under the direction of F. N. Ratcliffe of the Wildlife Survey Section of the Commonwealth Scientific and Industrial Research Organization in the River Murray basin in 1950.

Recent Trials

Between May and November 1950 the virus was liberated at several locations near Corowa in New South Wales and Gunbower in Victoria (see Fig. 1). The disease did not appear to have become established, but during December myxomatosis was reported for 30 miles along both sides of the River Murray near Corowa. The disease then spread over a vast area with amazing

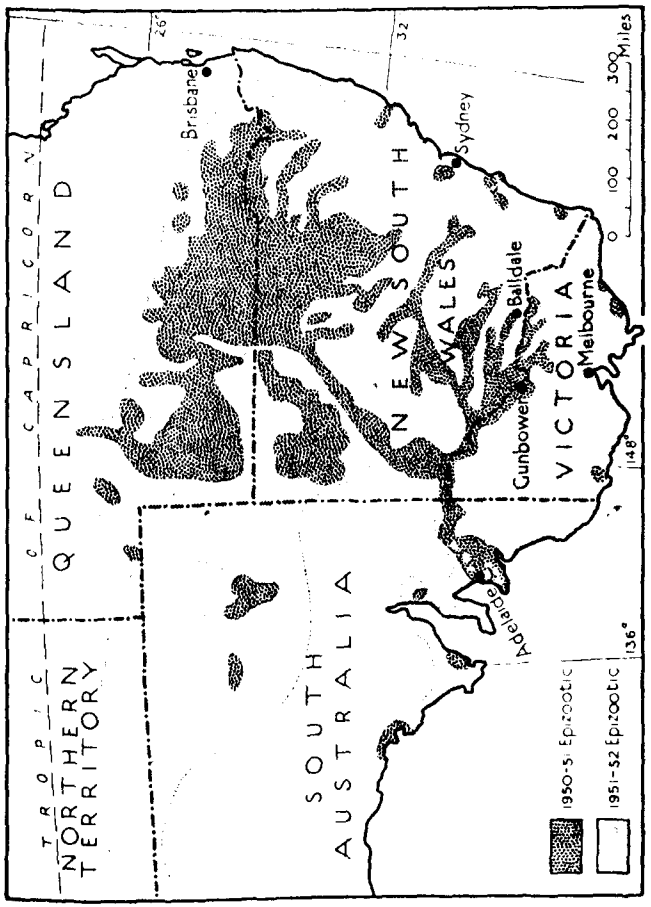


Fig. 1. Approximate areas over which myxomatosis epizootics occurred. The virus was first liberated at Gunbower and Balldale. Note proximity to rivers in the first epizootic, and the widespread character of the second, which extended between the rivers as well as over most of the areas of the first epizootic. Boundaries are only approximate, for information is very inadequate over large sections of the country, particularly towards the west and northwest. Most of the data were supplied through the kindness of Mr F. N. Ratcliffe. Advice on the New South Wales outbreaks was given by Dr Grahame Edgar; data from South Australia were from Mr E. W. L. Lines

rapidity, and by mid-February covered roughly the area indicated by shading in Figure 1. The disease was still spreading in April, but by the beginning of June the vast spread was considered to have ceased. Nevertheless cases of the disease continued to occur in small numbers in many localities. During the spring and early summer months of 1951, virus had been introduced artificially in rabbits in many areas of the States of South Australia, Victoria, and New South Wales. This artificial introduction of the virus, plus the natural spread during the winter which we have just mentioned, provided a multitude of foci which initiated highly destructive outbreaks, or 'epizootics' as we shall now call them, in many parts of south-eastern Australia. Many of these outbreaks occurred well away from watercourses, and both field and laboratory observations suggest that the most important mosquito vector differed from that which caused the violent epizootics of 1950–51.

An unexpected occurrence of the disease, which reproduced Sanarelli's original experience in Montevideo, was in the animal houses of the Baker Institute for Medical Research in the heart of the city of Melbourne. For a time in January 1952 it looked as though the entire stock of valuable experimental rabbits might be lost, but the disease was halted by slaughter of the affected animals and vaccination of the remainder with a vaccine prepared from the benign but closely related fibroma virus.

Method of Spread

So far we have considered where the disease was reported at a particular time, and not how it got there. Taking the most conservative view it will be seen that the disease was spreading along the Murray River in December 1950 and had spread well into Queensland to about the 26th parallel by February 1951 – a linear distance of some 700 miles within two months of its first recognition away from the experimental sites in Victoria. The spread was not continuous, but the disease was said to be present on practically every watercourse of any size over the area. How could this spread have occurred? The sparse population and the lack of knowledge of the disease rule out with practical certainty

the possibility of human intervention in the initial spread of myxomatosis, although, of course, landowners and the State Departments of Agriculture made every effort to stimulate the spread after the capabilities of the disease had been demonstrated. The route taken by the disease is away from any air lanes. Rabbits migrate sometimes, under conditions of poor food supply, for considerable distances, but migration could not have accounted for the extent of the spread of the disease. We are forced to consider the possibility that a mobile insect vector carried myxomatosis to the rabbits over the greater part of eastern Australia west of the Great Dividing Range, the mountain chain that runs north and south the length of the continent close to the eastern seaboard. The proximity of the disease to watercourses suggests the activity of a winged insect vector with an aquatic stage, and points to the mosquito. But movements of mosquitoes over distances of hundreds of miles have never been recorded. Indeed it would be difficult to devise an experiment to prove such a migration except by the introduction of a new disease which multiplied as it spread from a known point of introduction.

There are, however, several factors which suggest that mosquitoes could have carried myxomatosis over the vast distances covered in the 1951 epizootic. Much of the area was subjected in November 1950 to the worst floods in the century. Mosquitoes bred in the receding waters in plague numbers and it seems certain that a vast population was left without sites for egg-laying. Further, this area in western New South Wales is very subject to thermal up-currents and, in view of recent work in England and in the United States on the large populations of other kinds of insects which can get into the upper atmosphere, mosquitoes may well have been transported over great distances by air currents under the conditions obtaining at the time. It will be difficult to learn exactly how the disease spread, for conditions similar to those in 1950–51 certainly do not occur frequently in the region involved. Furthermore, the disease is now established in every State in Australia, and a repetition of spread from a single focus cannot be expected to occur again.

REASONS FOR THE EPIZOOTIC SPREAD OF MYXOMATOSIS

Earlier in our discussion we have briefly considered two of the principal characters in the disease, the virus and its host, and we have seen that myxomatosis in the European rabbits is unique in its very high lethality. Some infectious agents are always fatal when inoculated by special routes into laboratory animals, but no other infectious disease exists, except perhaps pneumonic plague in man, which has such a high mortality rate when spread by natural means. From the point of view of spread, what are the consequences of this lethality? The most important is that there are few survivors which might build up a rabbit population which is either protected by infection of the young while they enjoy the protection of antibody derived from the mother, or exhibits a higher innate resistance than the ordinary run of rabbits. If the virus were to maintain an almost 100 per cent lethality for the rabbit, it would be unlikely that a resistant population would be built up, and virtually every rabbit infected would die.

We have seen from the history of the preliminary field trials that myxomatosis did not previously spread as far and wide as the 1950–51 epizootic indicated. What was responsible for the difference between the experiences in arid north-eastern South Australia and the Murray Valley? We have outlined in the previous section the epidemiological considerations which indicated that another important character, or group of characters, entered the stage, namely mobile insect vectors, principally mosquitoes and sandflies.

Quite a number of virus diseases are normally spread by mosquitoes or other flying insects, and these diseases show a geographic localization which depends, to a large extent, on the distribution of the insect vectors. Could any such localization be made out in this outbreak? It is apparent from the description of the outbreak just given that in 1950–51 there was some measure of localization to the neighbourhood of watercourses, where mosquitoes could breed. But exceptions occurred, and the

distribution of myxomatosis did not correspond to that of any particular species of mosquito. The explanation for this became apparent when laboratory investigations were made into the mechanism of transmission of myxomatosis by mosquitoes. It was found that, unlike yellow fever, dengue, encephalitis, and so on, the transmission of myxomatosis, at least by the species used for the tests, *Aedes aegypti*, was a mechanical process. The mosquito could be thought of simply as the carrier of a flying pin, which became contaminated with virus by insertion through the virus-laden skin of an infected rabbit, and deposited the virus when another blood meal was taken from a normal rabbit. It was found that mosquitoes could do this repeatedly, and for many days. And the important consequence follows that *any* mosquito, or indeed any biting insect, that will feed on two rabbits can transfer the disease. Investigation by Bull and Mules, and by Aragão, showed that at least four Australian and two South American species of mosquito, as well as *Aedes aegypti*, could transfer the disease mechanically. The more frequently and the more exclusively a species of mosquito feeds on rabbits, the more efficient will it be as a vector.

The most important factor in determining the phenomenal *spread* of myxomatosis (but not its lethality, which is the other requirement for its effectiveness for rabbit control) is therefore not the virus nor the host, but the peculiarities of its mechanism of transmission by mosquito and other biting insects, such as sandflies.

In summary, the factors that account for the successful spread of myxomatosis in Australia are (1) the introduction of the disease into a large, completely susceptible population, (2) the extreme virulence of the causative agent and (3) the effective transmission by a variety of vectors at least some of which possessed extraordinary means of dispersal.

THE FUTURE DEVELOPMENT OF MYXOMATOSIS IN AUSTRALIA

When a pathogenic microbe is introduced into a completely susceptible population it spreads rapidly through it. We may

remember influenza in the great pandemic of 1918–19, and measles in the Fiji Islands. After a time which varies from a few to many years an equilibrium is established between parasite and host, so that both continue to survive. Will this happen with myxomatosis in Australia? If we look at the position in South America, the ancestral home of the virus, it is apparent that such an equilibrium has been established between the virus and the Brazilian native rabbit. But as we mentioned earlier, the position of myxomatosis in the European rabbit is unique. In the instances of human diseases mentioned above, the death rate did not exceed about 5 to 10 per cent of the population, and the survivors were immune to re-infection for a few years, and sometimes for life. Gradually, also, genetic selection for resistance helped to remould the population, an effect particularly noticeable in tuberculosis. It is too early to say whether establishment of equilibrium by this means can be expected in myxomatosis. The percentage of survivors (probably under 1 per cent) is very low, and the few experiments so far performed indicate that neither acquired immunity of a significant degree, nor enhanced inherent resistance to the virus, is passed from these survivors to their offspring. However, small resistant populations, if they did exist, would probably increase rapidly.

Another possible means to the establishment of equilibrium exists, namely a mutation in the virus so that it loses its extreme virulence for the domestic rabbit. We have laboratory evidence that such mutations can occur. Dr E. Weston Hurst some years ago developed a mutant strain of myxoma virus by serial brain-to-brain passage of the virus in rabbits. This strain killed only ten out of thirty rabbits inoculated in the skin, a mortality rate which immediately removes the disease from its unique position. Another investigator in America discovered a mutant form which caused disease with a mortality rate of only 15 per cent. In the tremendous number of virus particles produced in an epizootic which numbers its victims by the million the chances are that similar mutations to lessened virulence will occur. Unless the mutant virus is transferred to another rabbit, and the chances are considerable that it will not be, it is of no importance. If it is transferred, its possible importance depends partly upon the local situation at the time, i.e. whether it is early or late in the season, whether the local epizootic is just commencing or declining, and whether the local insect vector is one likely to move considerable distances or to remain in one locality.

In the course of years, however, we may imagine that such a mutation might occur several times, eventually in an environment favouring its dissemination. What happens then depends upon the relative efficiency with which the original virulent virus and the mutant form are spread. If the mutant with diminished virulence produces many and large skin tumours, which harbour virus in considerable quantity for a long time, it would be a more effective source of infection for the insect vectors than the original disease, which kills a big proportion of its victims before they develop any skin swellings whatever.

The outcome is unpredictable. It is possible that no such mutant virus will arise which is both less virulent and more infectious than the original form. And the natural situation may be modified by the systematic and widespread introduction by man of fully virulent virus into the rabbit population each year. It may even be possible to isolate, in the laboratory, a mutant which combines with its original virulence an ability to withstand the effects of protective antibody, and successfully infect and perhaps kill even the 'immune' rabbit. Something of this sort has already been achieved with influenza virus. We cannot yet predict whether a relatively resistant population of rabbits will emerge from the few survivors of past epizootics.

The future of myxomatosis as a weapon for rabbit control thus depends upon the maintenance or otherwise of its present unique virulence for the wild rabbits of Australia. Assuming that this is maintained, what is the position? It is not possible to define it accurately. First it is virtually certain that it will not eradicate rabbits from the continent. The area of Australia is too vast, the areas with many rabbits and few insect vectors too numerous and the migratory habits of the rabbit too well developed for this to occur. But, unless or until resistant strains of rabbit develop, kills of 90 per cent and over of the rabbits in

many areas can be expected every season that the weather conditions are suitable for the production of a large population of insect vectors. In many valuable pastoral areas this means practically every year, and in many other parts of the country exceptional seasons may produce mosquitoes in pest proportions, and a correspondingly violent epizootic of myxomatosis. Undoubtedly a new weapon of great importance has been added to the armamentarium of those concerned with rabbit control, but a repetition of the control of prickly pear by *Cactoblastis* is not to be expected.

FURTHER READING

The natural history of myxomatosis in South America is described in:

Aragão, H. B. *Memorias do Instituto Oswaldo Cruz*, **38**, 93 (1943).

For the technical aspects of the myxomatosis epizootics in Australia, *see*:

Ratcliffe, F. N., Myers, K., Fennessy, B. V., and Calaby, J. H. *Nature*, **170**, 7 (1952).

Studies in mosquito transmission are reported in:

Fenner, F., Day, M. F., and Woodrofe, G. M. *Australian Journal of Experimental Biology*, **30**, 139 (1952).

A more detailed account of the virus, and the immunological responses of the rabbit is given in:

Fenner, F. *Australian Journal of Science*, **15**, 81 (1952).

THE ORIGIN OF PETROLEUM

PAUL V. SMITH

The following is an extensive abstract of a paper presented on 30 December 1952 during the Annual Meeting of the American Association for the Advancement of Science.

DURING the last hundred years many hundreds of papers have been published on various aspects of 'the origin of petroleum'. Almost as many theories as authors have emerged from this mass of literature simply because of the elusive nature of petroleum itself. Coal, a much more co-operative subject for research, has yielded the secret of its origin because of its immobility. Petroleum, on the other hand, is capable of movement, and it has never been possible to say with certainty, 'This is where petroleum was formed'. Perhaps the most widely held premiss embraces the deposition of organic matter in a marine or brackish-water environment. Many have concluded that the action of heat and pressure over aeons of time is necessary to convert this organic matter into petroleum. Recently, however, a school of thought has developed which suggests that oil-formation might begin soon after the deposition of the organic matter in the sediments. Interest in this possibility was increased by the isolation of hydrocarbons from kelp and algae by T. S. Oakwood during 1944–5, and in 1947 a sub-committee of the American Petroleum Institute suggested that a study of the organic constituents of recent sediments would give useful information. Such an investigation was begun by the Chemical Division of the Esso Laboratories in 1950, when an opportunity arose to collaborate with geologists of the Humble Oil and Refining Company in the collection of samples.

Because earlier work had been done mostly on near-surface samples, it was decided to begin with a 'core' of recent sediments

extending to some considerable depth below the water-sediment interface. A site along the coast of the Gulf of Mexico seemed reasonable in relation to the assumed conditions of petroleum-formation, and, when drilling was started on a well some seven miles off Grande Isle, Louisiana, special attention was given to the collection of a core starting from the Gulf floor and extending to a depth of 106 feet below it. For comparison with different types of environment, a small core was collected in the Laguna Madre area of the Texas coast, and three sections of a core from the Santa Cruz Basin off the California coast were supplied by K. O. Emery and S. C. Rittenberg of the University of Southern California. A number of miscellaneous surface samples was also collected.

Attention was concentrated in the first place on that part of the material which can be extracted from the sediments by boiling samples from different depths with a mixture of benzene, methanol and acetone. Chromatographic analysis,* using powdered activated alumina, was at once tried as a means of separating the material thus extracted. The procedures followed were already in use in the laboratory on other problems. The samples were introduced with *n*-heptane at the top of the alumina-filled column, and the column was then successively washed (eluted) with *n*-heptane, benzene, pyridine, acetone, and methanol. Under these conditions, paraffins and naphthenes are removed from the column by *n*-heptane, aromatic compounds by benzene, and asphaltic or oxygenated compounds by pyridine, acetone, and methanol. In view of earlier statements that liquid hydrocarbons had never been identified in recent sediments, it was gratifying to detect liquid aliphatic and aromatic hydrocarbons in our first experiments. Some of the data for the Grande Isle core are summarized in Table I below. It can be seen that the amount of free hydrocarbons (paraffin-naphthene plus aromatic fractions) increases with increasing depth of burial, whereas the amount of asphaltic material and complex organic matter (asphaltic fraction, plus material remaining on

*See article by E. Lester Smith in *Science News 13*; or, for a recent short reference, *Science News 26*, 88 (1952).

TABLE I: *Variation with Depth in the Proportions of Hydrocarbons and other Constituents in dissolved extracts from 106-ft Grande Isle Core*

Depth (feet)	Free Hydrocarbons (per cent)	Asphaltic and complex organic matter (per cent)
3–4	7.5	92.5
18–22	20.4	79.6
52–53	27.5	72.5
102–103	30.9	69.1

the column) decreases with depth. The total quantity of organic matter extracted was about the same in these four sections, a little over 3 parts in 10,000 dry weight in each case. From this it can be inferred that the organic content is being changed in the direction of petroleum with increased depth of burial or greater age.

Support for this conclusion was obtained by other methods. Normal chemical analysis showed an increase with depth in the hydrogen/carbon ratio; and, when samples of dried sediments from different sections were heated in an air-free steel bomb at 450° C., the quantity of hydrocarbons found in the gaseous products was again greater with increasing depth of sediment.

Although extreme care had been taken to avoid contamination of the samples between collection and analysis, it was conceivable that the detected hydrocarbons might have entered the sediments by seepage, and that they might be of ancient origin. However, thanks to the development of the carbon-14 method* of determining the age of organic matter up to 30,000 years, it was possible to settle this point unequivocally.

Arrangements were made with Dr J. Lawrence Kulp, of the Lamont Geological Observatory, Columbia University, to measure the carbon-14 content of several samples. To obtain big enough samples of hydrocarbons, the available core sections were combined into two portions and extracted in a pilot-plant

*Notes on this method have been published in *Science News 10*, 28 (1949) and *Science News 17*, 115 (1950).

which had been assembled for such work. Even so, the usual procedures had to be modified to enable carbon-14 measurements to be made on samples containing from 0.75 to 1.2 grammes weight of carbon instead of the usual 8 to 10 grammes weight.

Apparent ages ranging from 11,800 to 14,600 years, plus or minus 1,400 years, were obtained for the hydrocarbons from sections of the Grande Isle core. A composite sample of carbonate from the entire core gave an apparent age of 12,300 years, plus or minus 1,200 years. For control purposes, similar measurements were made on extracts from sediments undoubtedly of ancient origin, and also on a hydrocarbon produced from sperm oil. The hydrocarbons from the ancient sediments were found to be 'dead', as might be expected; that from sperm oil agreed with measurements made on modern wood under the same conditions.

Sufficient material was not available to determine whether hydrocarbons were present in the first, and 'youngest', foot of sediments in the Grande Isle core. A larger sample was therefore scooped from the floor of the Gulf off Corpus Christi, Texas. Extracts from this sample differed little either in total organic matter or in analysis from the section of the Grande Isle core 3–4 feet below the surface.

Having thus established the presence of hydrocarbons in samples from the floor of the open gulf, it was of interest to examine sediment samples from an entirely different type of environment. Such samples were available from the Laguna Madre area of the Texas coast. This is an area of which portions are alternately submerged and slightly out of water. Pools of salt water form in many places and gradually evaporate to produce a super-saline environment. Algae are quite abundant. Again, hydrocarbons have been detected in the organic matter extracted from these sediments, although the type of breakdown is different. Aromatics are either missing, or are present only in very small quantities, in contrast with the Grande Isle sediments in which appreciable quantities were found. Also, that part of the organic content which is extractable with solvents became

greater with increasing depth in the Laguna Madre core. This must mean that those organic molecules which can be solvent-extracted are more resistant to breakdown, whether by bacteria or through chemical changes.

There were also differences in the sections provided by Emery and Rittenberg from the Santa Cruz Basin of the California coast. Here the extractable organic content is at least three to four times as great as in the Grande Isle core, and more aromatics than paraffin-naphthenes were found in all three sections. This is the reverse of the situation in the Gulf coast, where all extracts were more paraffinic.

ORIGIN OF THE HYDROCARBONS

It is interesting to speculate as to the source of the hydrocarbons found in the sediments. In a bibliography compiled several years ago there were listed 74 references to the reported isolation of hydrocarbons from lichens, algae, higher plants, insects, worms, fishes, and higher animals including man. More recent work by C. E. ZoBell and his colleagues has shown that aliphatic hydrocarbons are present in the cell structure of certain bacteria – sulphate-reducing species, which can be cultivated in an inorganic medium. It is therefore of possible interest that in two shallow-depth samples from different sites in the same area (Laguna Madre), there is an apparent correlation between the amounts of sulphur and of free hydrocarbons found in the samples. It is also conceivable that most bacteria may contain hydrocarbons in minute quantities. However, it has been shown by ZoBell that the bacterial population falls off rapidly below the first few centimetres of surface, and in a recent extension of this work we have found that the numbers of bacteria become insignificant with increasing depth. Any contribution by bacteria must therefore be limited to the area within comparatively few feet of the surface.

Samples of bluefish and of oysters have also been examined in our laboratory, using the same methods that were applied to the sediments. Fifty-eight parts per million of the dry weight of the bluefish consisted of paraffin-naphthenes, and there was also

a blue-fluorescent, aromatic-containing fraction. The oysters yielded a paraffin-naphthene fraction amounting to 45 parts per million of their dry weight.

In spite of the many references quoted in the bibliography already mentioned, the number of organisms which have been examined for hydrocarbons is quite small, and most of the observations have been made during researches undertaken with other objects. It is unfortunate that so little is known about the hydrocarbon content of marine organisms in particular. With this reservation, it appears that hydrocarbons of various types are present in many forms of plant and animal life, and the work done to date suggests that the phenomenon may be quite general. Further, it is probable that such materials are more resistant to breakdown than are most types of organic matter, and hence would tend to accumulate in the sediments relative to their original concentration in the organisms.

It seems plausible, therefore, to suggest that petroleum is being formed in the present era, and that the crude product is nature's composite of the hydrocarbon-remains of many forms of marine life. Changes in the relative contributions of these different organisms would lead, obviously, to crude oils of varying composition. Much more work will, however, be needed to clarify the picture.

There remains for consideration the question of quantity. Extrapolation of the data obtained for the Grande Isle core gives a figure of 4,500,000 barrels of oil per cubic mile of sediments, or 7,000 barrels per acre for a column one mile deep. A similar extrapolation of the Santa Cruz Basin sediments would give slightly greater figures. It would seem, therefore, from work to date that there are enough hydrocarbons in recent marine sediments to account for respectable oil pools if they could collect in some favourable porous stratum. The problem of how these small, highly dispersed droplets in the clays collect to give an oil accumulation is important and is being investigated.

THEORY OF GAMES

S. VAJDA

FROM a variety of considerations which, not very long ago, would have seemed in no way to be connected, a new branch of science has emerged: the Theory of Games. The origins of this theory go back to 1928, when John von Neumann read a paper to the Mathematical Society of Göttingen, introducing new and unorthodox concepts. During the last war Professors von Neumann and Morgenstern published their monumental and painstaking treatise *Theory of Games and Economic Behavior*, and scientists concerned with operational research on both sides of the Atlantic investigated and developed various theoretical aspects of tactics and strategy. The theory does not help one to become proficient in any specific games. It does not tabulate chess openings or advise on poker biddings. It is concerned, not with any particular game, but rather with general aspects applicable to all games, and with processes which obtain a special significance when a long succession of plays is being considered. Such conclusions as that you cannot win at Monte Carlo in the long run may emerge as a trivial consequence of the theory, but statistics is not of its essence. Being a mathematical theory, it draws on the results of several branches of mathematics, such as algebra and measure theory, but entirely new concepts had to be created as well, and it is in these that the particular attraction of the subject lies.

The essential feature of any game is the fact that one has to do with one or more opponents and that, therefore, only some of the relevant variables are under the control of any single player. It is clear that a theory which takes account of this peculiarity can be applied to the analysis of warfare, to economic problems, and even to decisions which must be taken where no specific opponent appears to exist, but where all variables outside one's

own control are dependent on 'Chance' or on 'Laws of Nature'. It is the aim of this article to introduce the reader to some of the new concepts and to show how they emerge naturally from a consideration of typical features of games.*

MATCHING PENNIES

To begin with, let us consider one of the simplest games imaginable, that of 'Matching Pennies'. Two players put down a penny each, either head or tail up, unknown to the opponent. They then uncover their coins and A receives his own and also B's penny, if both coins show the same side. Otherwise B collects both pennies. Clearly it is A's aim to show the same side of the penny as B does, and all his decisions connected with the game depend on this aim of his and on his ignorance of B's procedure.

It is convenient to introduce here a few definitions. There are two persons to play Matching Pennies and one person wins what the other loses. In other words, the gains and losses of the two players are balanced. Such a game is called a zero-sum two-person game. By a 'game' we understand in this context the aggregate of rules which set out the possible behaviour of the players and their gains and losses. Thus 'game' is an abstract concept. One particular instance of it, as it is actually performed, is called a 'play', consisting of a set of 'moves'. Matching Pennies exhibits an extreme simplicity because it finishes after only one move of each player.

In more complicated games one imagines, as a rule, that the players decide on their own moves as the play develops. This is, however, not essential. One could equally well imagine that each player decides before the play begins what 'strategy' he wants to apply. Such a strategy determines in advance what moves the player would make under all conceivable circumstances. One could imagine that each player chooses his strategy unknown to the other player and informs an umpire of his choice. The latter would then consult a list containing all possible pairs of strategies of the two players and would read out the result. It

*The case of 'noughts and crosses' was discussed by D. W. Davies in *Science News 16.*

would not be necessary to play any more. (Compare D. W. Davies's 'foregone conclusion' in his article already cited.) Fortunately the number of possible strategies in such sophisticated games as, for instance, chess is so enormously large that there is no danger that their complete enumeration and evaluation will ever kill the interest in the game.

Reverting to Matching Pennies, we can construct a table showing the payments which A receives and how they depend on the possible outcomes of a play. Such a table is referred to as the 'pay-off table' (for A) and looks as follows:

TABLE I: *Pay-off Table for Matching Pennies*

		B's coin	
		Head	Tail
A's coin	Head	1	−1
	Tail	−1	1

It is clearly impossible for either player to choose his move or his strategy (the two concepts are equivalent in this particular game) in such a way that he can be sure of winning. On the other hand, any player who knows the opponent's move can win. It would be very bad policy to stick to any particular move in a succession of games, since this would soon be noticed. It is therefore natural to suggest that a player should change his strategies at random, i.e. in a manner that does not give any indication of the strategy to be used in the next play. He can still decide on the over-all proportions in which his strategies should be used in the long run. It is, of course, also possible that the opponent finds out what these proportions are. Each player should therefore choose them in such a way that his position does not become worse when he is found out. It is clear that if A chooses his originally given 'pure strategies' in the proportion $\frac{1}{2}:\frac{1}{2}$, then, whatever B does, A's gains will equal his losses in the long run. But this is, in fact, the best A can hope to achieve. If he chooses his strategies in any other proportion, say once head to three times tail, then B, provided he knew it, would be shrewd enough

to choose always head, because then A would lose more often in the long run than he would win. Reversing the argument, it is seen that B should, in ignorance of A's strategy, once again choose head and tail with equal frequencies, in order to view without concern the danger of being found out. (This principle has been described as expecting the best in the worst possible world. We leave it to the reader's temperament to decide whether this description, if justified, should induce the player to gamble excessively rather than to play safe.) If, then, both players use their pure strategies in the proportion $\frac{1}{2}:\frac{1}{2}$, a certain stability will be reached in the sense that no player will change his behaviour, even if he finds out what his adversary does, since by doing so he would not improve his prospects. The pair of frequencies, which leads to such a stability, is called a 'solution' and the decision of a player to use his (pure) strategies in given proportions is itself a strategy and is called 'mixed'. We have just seen that in Matching Pennies no pure, but only a mixed, strategy leads to a solution. However, this is by no means always so, as can be shown by considering a simple modification of the game.

MODIFIED MATCHING PENNIES

We assume now that any player is allowed to 'call off' the play and that he receives, if he is the only one to do this, a halfpenny from his opponent, whereas no payment is made if both players call off. The corresponding pay-off table (again for A) is shown as Table II.

TABLE II: *Pay-off table for Modified Matching Pennies*

		B's coin			
		Head	Tail	Call off	Row minima
A's coin	Head	1	−1	$-\frac{1}{2}$	−1
	Tail	−1	1	$-\frac{1}{2}$	−1
	Call off	$\frac{1}{2}$	$\frac{1}{2}$	0	0
Column maxima		1	1	0	

An inspection of the table shows that now both players can choose pure strategies which give a solution. To achieve this, they must both call off. In any other case the player who finds out his opponent's intention can win a penny.

The entry 0, corresponding to the pair of strategies which form the solution, is called the value of the game. It is the smallest number in its row and the highest in its column. Therefore its position is called a 'saddle point', by an obvious analogy. It can easily be proved that if there are more saddle points in a pay-off table, then the numbers in all of them must be equal. A simple way of finding a saddle point, if one exists, is to write down the column maxima and the row minima, as we have done in the margins of the table above. In this example the maximum of the row minima equals the minimum of the column maxima, viz. 0 (the value of the game). In more abbreviated terms, the 'maximin' equals the 'minimax'. The common value indicates a saddle point, and the corresponding solution is called a 'Minimax Solution'.

Now one can easily prove that the maximin can never exceed the minimax, but as we have already seen (see Matching Pennies), the two need not be equal. However, it has been proved that the maximin of all possible strategies, including mixed ones, is always equal to the minimax of all possible strategies of the opponent, again including mixed ones. Once more, the resulting pay-off is called the value of the game. The equality of the maximin and the minimax, extended to refer to mixed strategies, was proved first by John von Neumann in 1928. He called it the Main Theorem. His proof has since been simplified in various ways, but it is still too involved to be included in this merely expository article.

THE THREE BOXES GAME

The game of Matching Pennies led to an obvious best mixed strategy and its modification had a Minimax Solution of pure strategies. We shall now deal with a game whose solution is again given by mixed strategies, but not by such obvious ones. Let three boxes marked 1, 2, and 3 be given, containing these num-

bers of shillings respectively. The 'banker' removes the bottom of one of the boxes, but this is not discernible from the outside. A 'player' puts into two of the boxes the amounts of shillings marked on them. He then receives all the money in those two boxes. He will, of course, lose the money that he happened to put into a box without bottom. The pay-off table for the player is easily constructed (Table III). For instance, if the bottom of box 3 is missing and if the player puts 1s and 2s into the appropriate boxes, then he wins 3s altogether. However, if he puts his money into boxes 1 and 3, then he wins 1s and loses 3s, so that, on balance, he suffers a loss of 2s. In this way the following table has been constructed:

TABLE III: *Pay-off table for the Three Boxes Game*

		Bottom removed from box			
		1	2	3	Row minima
Money put into boxes	1, 2	1	−1	3	−1
	1, 3	2	4	−2	−2
	2, 3	5	1	−1	−1
Column maxima		5	4	3	

This table has no saddle point, since the maximum of the row minima is -1, whereas the minimum of the column maxima is 3. Therefore the solution (which is known to exist by virtue of the Main Theorem) must contain mixed strategies.

In order to find the solution, one could first try to see whether three positive numbers a, b, and c, adding up to unity, can be found such that every row gives the same value if the banker follows a mixed strategy in which the bottoms of the three boxes are removed in those proportions. This is the condition that the player will gain no benefit from discovering the strategy used by the banker. We could then repeat the procedure for every column, and hope in that way to obtain the player's mixed strategy such that the banker gains nothing from its discovery.

In the present case, the conditions for the banker's mixed strategy are (for the rows) that $a+b+c=1$, $a-b+3c=2a+4b-2c=5a+b-c$. If the banker chooses the proportions $a=5/22$, $b=8/22$, and $c=9/22$, which are consistent with these conditions, then the player wins 12/11, whatever he does, whereas if the banker chose any other proportions, the player could win more than this amount. By the same argument, we must try to find three numbers x, y, and z, say, so that $x+y+z=1$ and (applying these proportions to the columns) $x+2y+5z=-x+4y+z=3x-2y-z$. However, this results in $x=7/11$, $y=5/11$, and $z=-1/11$. Since z is negative, this is not a possible frequency, and we must find some other method for finding a solution. It would be too long to explain here how this is done systematically. The solution turns out to be (3/5, 2/5, 0) for the player and (0, 1/2, 1/2) for the banker.

When the banker uses the proportions (0, 1/2, 1/2), the gains of the player for his pure strategies are $-1/2+3/2=1$, $4/2-2/2=1$, and $1/2-1/2=0$ respectively. The latter will, therefore, be wise to use mixed strategies, in which only two pure strategies are combined—those in which he puts money into the 1s and 2s, and the 1s and 3s boxes respectively—since by using the 2s, 3s strategy as well he would invariably reduce his expected average gain. The proportions in which he should use those strategies are 3:2, as already stated. Doing this, he gains 1 and the banker loses 1. (We may assume that the latter will demand a fee amounting to 1, the value of the game, for participation.) The player cannot improve on his gain of 1, as long as the banker adheres to the proportions (0, 1/2, 1/2) on his part. For any other choice of the banker's, the player could gain more. Thus, for instance, if the banker used (0, 1/3, 2/3), the player could obtain a gain of 5/3, by choosing the pure strategy (1s, 2s). Conversely, if the player chose proportions different from those given for him, he could be made to gain less than 1, if he were found out. Thus, for instance, if he chose (4/5, 1/5, 0), then the banker might choose pure strategy 2, and the gain of the player would thereby be reduced, in the long run, to zero.

So far we have always assumed that all proceedings depend only on the decisions of the players. However, this would not

cover all possible games. We imagine now a slightly modified Three Boxes Game, where the banker does not decide on his own moves, but where they are chosen for him by a chance mechanism, which makes every pure strategy equally likely. In other words, his mixed strategy is (1/3, 1/3, 1/3), and this is known to the player, who must choose the pure strategy of always putting his money in the 2s and 3s boxes to gain in the proportion 5 : 3 over a series of plays; every other choice would give him less in the long run. This is an example of a game with so-called 'chance moves'. Obviously, these can also occur in combination with deliberate moves, and they may be available to both players. Card games, for instance, contain always both types of moves, because at least some of the cards are dealt at random (or should be).

In the Three Boxes Game both participants make their choice without knowledge of the other's moves, as was also the case in Matching Pennies, where indeed otherwise the game would have been senseless. However, games exist where the moves occur one at a time alternately and where, whenever a move is about to be made, all players know all previous moves, whether their own or those of the opponent (though not, of course, the opponent's strategy, which includes also all his future intentions for any possible contingency). Such games are said to have 'perfect information', whereas Matching Pennies, for instance, is a game with 'imperfect information'. It has been proved by von Neumann that a game with perfect information has always a saddle point and that, in a probability sense, this is also true of games with chance moves (and perfect information). With regard to games without chance moves, the theorem has, in effect, also been proved by D. W. Davies in his article in *Science News 16*, mentioned above. We shall now introduce a 'Modified Three Boxes Game' with perfect information which will, therefore, have a solution consisting of pure strategies for both players.

MODIFIED THREE BOXES GAME

The rules of the Modified Three Boxes Game are as follows: (1) the player puts his money into one of the boxes, (2) the banker, who has watched him, removes the bottom of one of the boxes, (3) the player puts his money into one of the two boxes not previously chosen.

The player has thus two moves. As to the banker, who has only one move, we can take it for granted that his best strategy is to remove the bottom of that box into which the player has put the money on his first move. Thus the move of the banker is known to the player in advance. The latter, however, has a choice of the following six strategies, with the gains as given (provided the banker uses his best strategy, as he will):

First move of the player	1	1	2	2	3	3
Second move of the player	2	3	1	3	1	2
Gain	1	2	−1	1	−2	−1

It emerges, then, that the player should first put 1s into box 1 (knowing that he will lose his money) and then 3s into box 3, obtaining an overall gain of 2s. This is, of course, also clear from common-sense considerations.

The Modified Three Boxes Game, besides being somewhat artificial, is also of rather restricted generality in that the banker has only one move and that, if this is known, his whole strategy is thereby known as well. But it will be clear by now that the same methods are applicable also to more complicated games. It is perhaps worth mentioning that there is no connexion between the division of games according to information and according to chance moves. This can be seen from Table IV:

TABLE IV: *The Division of Games on the basis of Information and of Chance Moves*

	Games with perfect information	Games with imperfect information
Games with chance moves	Backgammon	Poker
Games without chance moves	Chess*	Matching Pennies

*It is assumed that the strategies of chess-players are chosen when it is known who plays White.

THE TWO GENERALS

So far, we have considered games with saddle points, or games where the solution consists of a single mixed strategy for both players. We shall now introduce a game which has more than one single solution. The example will also illustrate how the theory of games may be applied to military considerations. We imagine the following simple strategic situation.

Two generals, A and B, face one another. A has three companies at his disposal and B has four. It is the aim of A to reach a town on one of two possible roads. B tries to prevent him from doing it. General A can send all three companies on the same road, or he may split up their number, but he may not divide up any single company. General B, again, may send all his companies on the same road, or on different roads, but he must keep any single company undivided. We assume that A has achieved his object if he has on any road a number of companies exceeding that of his adversary. If he reaches the town, he wins (indicated by 1), if he does not, the issue remains undecided (indicated by 0). The pay-off table (for A) is shown here as Table V:

TABLE V: *Pay-off table for the 'Two Generals' Game*

		Number of companies of B on the two roads, respectively				
		0 and 4	1 and 3	2 and 2	3 and 1	4 and 0
Number of companies of A on the two roads, respectively	0 and 3	0	0	1	1	1
	1 and 2	1	0	0	1	1
	2 and 1	1	1	0	0	1
	3 and 0	1	1	1	0	0

The row minima are all 0, the column maxima are all 1. Hence this table has no saddle point.

It is now possible to see, by inspection, that if A uses strategies mixed in any of the following proportions: $S_1=(\frac{1}{2}, 0, \frac{1}{2}, 0)$, or $S_2=(\frac{1}{2}, 0, 0, \frac{1}{2})$, or $S_3=(0, \frac{1}{2}, 0, \frac{1}{2})$, he will gain on the average at least $\frac{1}{2}$ in the long run, and that B can keep him down to this if he uses strategy $(0, \frac{1}{2}, 0, \frac{1}{2}, 0)=T$, say, but cannot reduce A's gain below $\frac{1}{2}$. On the other hand, if B uses strategy T, then A cannot obtain more than $\frac{1}{2}$, whatever he does. This means that any of the following pairs of mixed strategies, (S_1, T), (S_2, T), and (S_3, T), is a solution. The value of the game is the same for all solutions (namely $\frac{1}{2}$), and it can be proved generally that whenever a game has more than one solution, the value is the same for all of them.

General A has the three 'best' strategies S_1, S_2, and S_3 at his disposal, and it is obvious that, if he uses any combination of these mixed strategies, the resulting mixed strategy will also form together with T for General B, a solution in the sense (as before) that neither can lose from the discovery by the other of their respective strategies. Thus if r, s, and t are three positive numbers which add up to 1, but are otherwise quite arbitrary, then A may use the following mixed strategy: $(r+s)/2, t/2, r/2, (s+t)/2$. For example, taking $r=s=t=1/3$, the following strategy emerges: (1/3, 1/6, 1/6, 1/3). Together with $(0, \frac{1}{2}, 0, \frac{1}{2}, 0)$ for B, we obtain again the value of the game, namely one-half.

MORE PERSON GAMES

Games in which more than two players take part have not been so intensively studied as two-person games, though the great book by J. von Neumann and O. Morgenstern contains chapters on this subject as well. Pay-off tables can again be constructed, but no theorem similar to the Main Theorem exists. An essentially new situation arises from the fact that two players can now combine in 'coalitions' and play against the remaining player or players. This is quite a common situation in real-life problems, and sometimes, as at an auction, the coalition may be agreed upon in advance between some of the players. Only if this condition is satisfied can the effects of coalitions be treated by the theory in its present form.

Coalitions can also arise spontaneously in many-person games, which ostensibly are all-against-all, at a stage when some

one individual is too evidently winning. Coalitions of the latter kind are unpredictable, and may be changed and re-formed, according to the mood of the players, as the play proceeds. They cannot be treated by the theory, as at present developed, and illustrate one type of limitation in its application to real-life problems.

FURTHER READING

The best general reference is the difficult, now classical book:

von Neumann, J., and Morgenstern, O. *Theory of Games and Economic Behavior*, Princeton University Press, 2nd ed., 1947.

IRREVERSIBLE CHANGES: NEW THERMODYNAMICS FROM OLD

R. O. DAVIES

MOST people tend to think of thermodynamics as being like the Lord Chamberlain – old, very respectable, and thoroughly dull. Certainly a subject which is represented in this series by articles called 'A Hundred Years of Thermodynamics'* and 'A Century of Joule-Thomson Experiments'† can make no claim to novelty. But a discipline which regularly produces quite unexpected results from simple manipulations can hardly be called dull. Nor can certain recent developments yet be called respectable. It is these recent developments which will be discussed here. They have already received honourable mention from Professor Born, who wrote in his 'catalogue' article in *Science News 19*: 'Recently, attempts have been made to generalize thermodynamics to processes of stationary flow ... which seem to be of great importance for the understanding of living organisms.' Professor Born quoted also some names and dates, beginning with L. Onsager, 1930. Professor Onsager was born in Sweden, now works at Yale University, and has made a number of notable theoretical contributions to the no-man's-land between physics and chemistry. As will appear later, there are other, and earlier, names which could have been mentioned. But if use and interest of results are taken as criteria, then it is correct that it was the work of Onsager which brought the subject to life, and to the position, eventually, when an article on it is thought necessary to *Science News*. As its title suggests, his work showed the utility of making an extension from the older thermodynamics,

*Butler, J. A. V. *Science News 18*, 21 (1950).

†Charnley, A. *Science News 23*, 93 (1952).

dealing with changes described as 'reversible', to deal also with 'irreversible' changes which are more common. It is indeed one of the curiosities of science that the older thermodynamics should have led to such useful results when the conditions in which it can be strictly applied are so limited. To appreciate this point, we must consider first what is meant by 'reversible'.

Three main ideas are comprised in the concept of 'reversibility'. There is first the literal meaning, namely, that a reversible process can operate in either the forward or the reverse direction (in time). Secondly, at each stage of the process, the system is supposed to be 'near enough' to an equilibrium state: 'near enough', that is, for arguments based on the assumption of equilibrium not to be vitiated in practice. Finally, the rate at which the process takes place is either very slow or very fast, relative to any irreversible changes which affect the problem. These three notions have been written in order of increasing vagueness: the last particularly (which is closely connected with the second) requires to be illustrated by an example. One is afforded by glass, the equilibrium state of which is a smooth continuation of the liquid state. (We neglect here the possibility of crystallization – 'devitrification' – which would require a separate discussion on similar lines.) In practice, the structure of the glass-forming material gets frozen-in during cooling and, because of the large viscosity, would take thousands of centuries at room temperatures to reach its equilibrium condition. This irreversible change is so slow compared with changes such as those caused by compression that the latter can be treated as reversible and the glass can be considered as if in an equilibrium state – provided, of course, that the compression is slow enough for mechanical or thermal irreversibility to be negligible. Irreversible thermodynamics is important precisely because few processes of practical interest satisfy this requirement of being 'very slow and/or very fast' with respect to all relevant irreversible changes.

ENGINEER'S THERMODYNAMICS

There are two distinct ways of approaching thermodynamics, which (for the sake of a label) we can call the engineer's method and the chemist's method. The engineer's method was developed first, is easier to teach, and – perhaps for these reasons – is more widely understood. It is associated with the names of Mayer, Joule, Carnot, Clausius, and Thomson. Its main devices are that famous but imaginary piece of apparatus called the Carnot reversible heat engine, and its more generally useful counterpart, the Carnot heat cycle. Their use rests on two assumptions, which are expressed as the First and Second Laws of Thermodynamics. The First Law is a statement of the energy-equivalence of heat (the conservation of energy), and by itself is of limited value. But, if one assumes in addition a single 'postulate of impotence' as a statement of the Second Law, it is possible to use these devices to draw some very general conclusions about the thermal and mechanical properties of *bodies with fixed chemical constitutions*. As particular examples, it is possible to calculate not only the ideal efficiency of a heat engine, working between specified temperatures, but also the effect of pressure on the boiling point of water (experienced by climbers) and the melting point of ice (used in skating). In illustration of breadth of scope, it may be stated that entropy – which will be defined in the following paragraph – is that quantity which increases when a heated object cools, when water runs from reservoir to kitchen sink, and when an electrical condenser discharges.

Perhaps the simplest 'postulate of impotence' is the one, due to Clausius, which says that it is impossible to construct a device which will transfer heat from a cold body to a hot body with no net change in the condition of the auxiliary apparatus. From this one can deduce two main results:

(*a*) There exist for bodies in equilibrium quantities S (the entropy) and T (the absolute temperature) which are fixed uniquely by the state, and which are connected by the fact that for small reversible changes of state the change in entropy is equal to the amount of heat entering the system divided by the absolute temperature: $dS=dQ/T$. Or, as more often stated: $dQ=T.dS$.

(*b*) When a completely isolated system (i.e. one in which all thermal and mechanical interactions take place within the sys-

tem itself) goes from one equilibrium state to another, the sum of the entropies of all bodies within the system can only increase.

From the first of these results there follows a whole host of useful (and sometimes surprising) physical relations. They are obtained by replacing the heat flow (dQ) by its energy equivalent. This is done by using the First Law of Thermodynamics – the conservation of energy – which tells us that the change in energy (dU) is equal to the heat entering the system (dQ) plus the mechanical work done on the system. Of the two latter quantities, dQ can be replaced, as we have just seen by $T.dS$; and the mechanical work can be expressed in terms of the pressure (p) and the change of volume (dV) – negative in compression – as minus $p.dV$. So we reach the relationship that the change of energy (dU) equals $T.dS - p.dV$. Finally one must use the fact that the energy (U) is fixed uniquely by the state.

An elementary but characteristic result obtained by the use of this method may be worth mentioning. It was found by experiment (before the discovery of the Second Law) that gases under 'ordinary' conditions have two important properties:

(1) The pressure, volume, and temperature of a gas in equilibrium are related in a manner most succinctly summarized by the equation $pV=RT$, in which R is a constant (Boyle's Law, Charles's Law).

(2) The energy (U) of a gas depends only on the temperature and is unaffected by changes in pressure or volume provided that the temperature is held constant (Joule's Law).

By using the Second Law in the way indicated above, one can show that if property (1) is assumed then property (2) can be deduced from it. The proof of this rests wholly on the Second Law of Thermodynamics; it cannot be recovered from the First Law alone. (It is interesting to notice that property (1) cannot be completely deduced from property (2).)

The second main result of engineer's thermodynamics – that the sum of the entropies of all bodies within a completely closed system can only increase – is of quite a different nature. At one time or another most of us have been puzzled by the question: why is it that when you look at yourself in a mirror your left

and right hands are interchanged, but you are not upside down or round about? On (metaphorical) reflexion we see that the explanation is connected with the fact that the notions 'up' and 'down' are defined by referring to a physical object, e.g. 'down' means 'towards the ground'. 'Back' and 'front' are similarly defined. But 'right' and 'left' are described with reference to the notions 'up' and 'front' already established. In an analogous way, if (instead of looking in a mirror) we interchange the words 'hot' and 'cold' in our postulate of impotence we find that proposition (b) is inverted – the entropy of an isolated system *decreases* – while proposition (a) remains unchanged. The analogy suggests – although of course it does not prove – that just as the notions of 'right' and 'left' would, if we were suddenly translated to the French Assembly, allow us to distinguish Communists from Gaullists before they spoke, so does proposition (b) allow us to decide which of two given possible states of an isolated system would precede the other in time. To do this we only have to find out which state has the greater entropy.

As Clausius recognized, this aspect of the Second Law (as distinct from the discussion of steam engines) is of great philosophical importance. There is as yet no general agreement on such topics as the role of convention in thermodynamics and the 'running down of the universe'. Stephen Toulmin has pointed out that the 'isolated system' statement given above is not necessarily applicable to the universe as a whole and that scientific popularizers have here misled the public by exaggerated claims.* This is undoubtedly true, but Toulmin himself has done scant justice to the more extended type of thermodynamics called chemical thermodynamics. This constitutes the chemist's approach to the subject, due in its essentials to Willard Gibbs, and it has been the starting point for more recent work.

CHEMIST'S THERMODYNAMICS

Gibbs saw that by some very plausible additions to the results given above it was possible to set up a scheme which was not limited to describing the thermal and mechanical properties of

**Listener 45*, p. 387 (March 8, 1951).

bodies of fixed chemical constitution. It provided also for changes in composition (due to substances going into or leaving solution and to chemical reactions), for electro-chemistry, and for the physical chemistry of surfaces. This development was analogous to that of other branches of science. Newton's dynamics were first applied to simple systems consisting of one or two particles and one or two rigid bodies. While the magnificent generalizations of Lagrange, Poisson, and Hamilton changed the face of the subject, they did not add any new fundamental law. So it was with Gibbs's theories.

His procedure can be illustrated by analogy. The output of an industry can be increased, in general, both by improved organization and by providing extra men and machines. Changes in organization correspond with the causes of change in energy that are studied in engineer's thermodynamics, and additions of men and machines with changes in composition. It is a reasonable first approach that, if small changes of both kinds are made at once, their effects will be additive.

Gibbs started by observing that the energy of a system must be altered when the composition is changed. Now it is at least plausible that, if the mass (M_1) of the first component of a mixture is altered by a small amount (dM_1), the change in energy *due to this cause alone* will be proportional to dM_1. Let us call the actual contribution $\mu_1.dM_1$, where μ_1 is a constant of proportionality. A similar argument can be applied to the second component, so that a change in mass of both components may be expected to yield a contribution of $\mu_1.dM_1+\mu_2.dM_2$ to the change in energy. This is the addition made by Gibbs to allow for changes in composition. The expression to which it must be added is the one already obtained from the engineer's thermodynamics for changes in energy, that is, $dU=T.dS-p.dV$. For a system whose composition can alter, the complete expression for the change in energy is therefore

$$dU=T.dS-p.dV+\mu_1.dM_1+\mu_2.dM_2+\ldots$$

This is 'the Gibbs equation'. It may be applied to any homogeneous part of a system, and is obviously important for dealing with systems that can exchange matter with their environment

('open systems'). Such systems have indeed been prominent in recent applications of thermodynamic methods in biology. The idea is to treat the biological unit – whether it be a macromolecule, a cell, or a complete living organism – as an open system. Irreversible changes are naturally very important in this work.

The coefficients μ_1 and μ_2 in the Gibbs equation are called the 'chemical potentials', and physical meanings can be given to them. This is done by imagining that a membrane, which allows only one component to pass through it, say component 1, is placed in between the system itself and a pure sample of this component. By closing off the combined arrangement (the system *and* the pure component) and applying the Second Law, one can prove that μ_1 is related to the physical condition of the pure sample when the whole is in equilibrium. This relation is $\mu_1=u-pv-Ts$, where the lower-case letters for u, v and s indicate that these quantities refer to unit mass, e.g. v is the volume per unit mass, or the inverse of density. The combination is often denoted by the symbol g and is called the Gibbs free energy per unit mass of the pure component in question. Notice that for a system of one component the Gibbs equation can be written as $dU=TdS-pdV+gdM$.

Not only does the Gibbs equation provide the new possibility of dealing with mixtures etc., but the use of it constitutes a great methodological advance on the older methods of cycles. Professor Guggenheim has justly remarked that while thermal cycles can always be used to solve problems, simple cycles are rarely accurate and accurate cycles are rarely simple. The key to the successful use of Gibbs's methods is the old physicist's dodge of thinking of entropy, energy, etc., as *things* or, rather, as substances spread through space and flowing hither and thither like water in a bath. However, if we attempt to carry this too far we run up against a difficulty. The case of energy is, if not completely satisfactory, at least tolerably so because the energy (like the water) is conserved. With entropy this is no longer true. Indeed the Second Law tells us that while the energy of an isolated system is fixed, its entropy increases as long as there are any changes going on inside it.

The most general aim of the new irreversible thermodynamics is to give a quantitative account of how the entropy increases. This means saying where it is produced and how much is produced under given circumstances. It is not always possible to do this, but when it is the Gibbs equation (or a suitable modification of it) must be used.

If all that was involved were this question of accounting for entropy production we could say, in the first place, that there is nothing very new about the idea. Many of the simpler cases of irreversibility such as heat conduction, electrical flow, and viscous flow were studied about the turn of the century by Bertrand, Duhem, Natanson, and Jouguet. Secondly, such a study would not be expected to produce any new and striking prediction. The comparative obscurity even of Duhem testifies to the fact that none was produced. (It is also a reflexion of the way that scientists demand novelty and originality of their great men. It is not enough to write [as Duhem did] a critical and systematic exposition of the whole of existing knowledge in some broad field.)

The question of entropy production and energy dissipation arose from the dead in the late 1930s following a discovery by Lars Onsager. As long as we restrict ourselves to considering only a single dissipative process, then our results, even when new, will be strictly in the spirit of classical thermodynamics. Onsager noticed that when two or more processes take place together a relationship exists between certain flow coefficients. This leads to new connexions between the properties of systems which, although not in equilibrium, are 'stationary' in the sense that the conditions of any part of the system remain constant.

ENTROPY PRODUCTION AND RECIPROCITY

Before describing Onsager's discovery, we must first examine how the entropy production can be found for simple processes. The essential steps are all present in a classical calculation which applies to heat conduction in a solid. Consider a solid slab (as in Figure 1) conducting heat between two opposite faces which are held at temperatures T and $T+\Delta T$. (ΔT is taken to be small with respect to T, but not necessarily infinitesimal.)

Assume for a start that we have stationary conditions. As we have just seen, this means that although an irreversible change is taking place, the temperature at any fixed point remains constant in time. The First Law now tells us that since the condition of the slab does not change as time goes on, the heat flowing out from one side of the slab is the same as the heat flowing into the other. Call the flow (from T to $T+\Delta T$) J_q per unit of time.* (J_q will be negative if ΔT is positive, since the heat flows from hot to cold.)

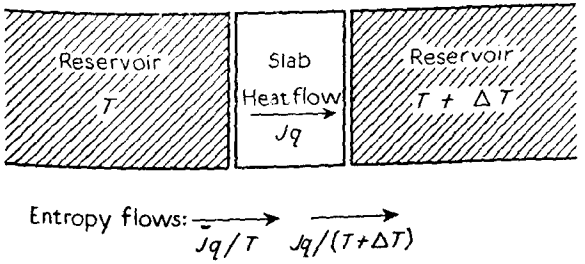


Fig. 1. To show the rate of production of entropy in a slab, the opposite sides of which are held at constant, but different, temperatures by two large reservoirs. The entropy flows into, and out from, the slab are shown in the figure. The rate of entropy production in the slab is $J_q/(T+\Delta T)-J_q/T$.

Now imagine that the temperatures are held fixed by two large reservoirs and consider the whole system (slab + reservoirs) as isolated (see Figure 1). To find the rate of increase of entropy we note first that the slab itself is in a steady condition so that its entropy remains unchanged. The rate of increase of entropy of the reservoirs is $-J_q/T+J_q/(T+\Delta T)=J_q\Delta(1/T)$. This quantity is therefore the rate at which the entropy of the whole system is increasing. The entropy must be considered as being *produced* within the conducting slab. Such an interpretation allows us to retain the 'thing-like' picture of entropy: there is an entropy

*From now on, capital J is used to denote a flow and a subscript to indicate the nature of the flow.

flow of J_q/T into the slab; an amount $J_q\Delta(1/T)$ is produced within the slab; and, since the entropy of the slab remains fixed, the flow out of the slab is $J_q/(T+\Delta T)$, which is seen to be the sum of the other two quantities just mentioned. We could, indeed, have found the entropy production by starting with the entropy flows to and from the reservoirs.

If conditions are not stationary (for example, before the final steady temperature-gradient in the slab has been established) we must take into account the entropy increase of the slab itself. The heat flows on either side of the slab will no longer be equal and their difference will give the rate of change of the energy of the slab. An application of the Gibbs equation then yields the entropy change in the slab *provided* that conditions are such that we can reasonably assign to the slab the mean temperature of the two reservoirs. The rate of entropy production again turns out to be $J_q\Delta(1/T)$ where J_q is now the mean heat flow.

Let us note the elements of this deduction which can be generalized to more complicated arrangements:

(*a*) Reversible reservoirs were provided at the points of the slab boundary where energy transfer took place. The essential role of these reservoirs was to establish the rate of entropy flow across the boundary.

(*b*) The principle of conservation of energy (the First Law) was used to find the rate of increase in energy of the slab.

(*c*) The Gibbs equation was used to find the entropy increase of the slab.

(*d*) The irreversible entropy production was found as the difference between the rate of increase of entropy of the slab and the net entropy flowing into it from the reservoirs.

To this list we may add a further principle which will be required presently:

(*e*) The conservation of mass enables us to express the rate of increase of mass in terms of the mass flows. (This is analogous to (*b*).)

There are two more lessons to be drawn from our simple result. The first is that the rate of entropy production ($J_q.\Delta(1/T)$) is of the form (flow of heat) × (difference in inverse temperature); or more briefly – since we may regard the difference as a 'force' tending to 'push' energy through the slab – rate of entropy production = flow × force. It turns out that for more complicated systems the rate of entropy production is the sum of several such expressions.

The second lesson concerns the relation between the heat-flow J_q and $\Delta(1/T)$. From the Second Law (increase of entropy) all that we can say is that the product $J_q.\Delta(1/T)$ is positive, so that J_q and $\Delta(1/T)$ must have the same sign. However, the law of heat conduction tells us that J_q is proportional to $-\Delta T$ so that, provided ΔT is not too large, we can write $J_q = k\Delta(1/T)$ in which k is proportional to the thermal conductivity of the specimen.

There are other well-known examples of linear relations between 'flows' and 'forces': the current in a conductor is proportional to the applied e.m.f. (Ohm's Law); diffusion flow is proportional to the concentration gradient (Fick's Law). The *assumption* of such relations is an essential part of the Onsager theory. Their range of validity is not determined by the theory, but must at present be found experimentally.

To find the entropy production for more complicated systems, we can start by asking the following questions: if mass and energy flow across a surface in a fluid, how much entropy is carried across the surface? Suppose that the dotted line in Figure 2 represents the surface, that the mass flows of the various constituents are $J_{m_1}, J_{m_2}, \ldots$ and that the energy flow is J_u. We are not concerned with what happens on the right of the surface. There may be, for example, a porous plug of some sort which cannot itself be described thermodynamically. On the left of the

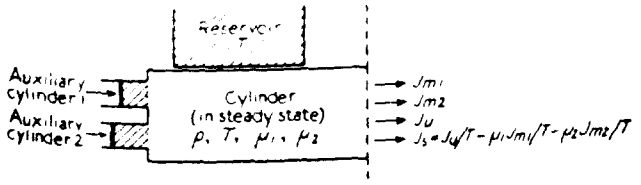


Fig. 2. The flow of mass, energy, and entropy across a surface.

surface we shall suppose that there is a large cylinder kept in a steady state by means of a heat reservoir and auxiliary cylinders which supply just enough of the various constituents to compensate for those drawn away across the dotted surface.

Everything on the left of the surface of interest can now be treated as reversible, that is, there is no entropy *produced* in this region. It follows that the entropy flow across our surface is just the rate at which the entropy of the cylinders and reservoir is decreasing. By using the Gibbs equation and the conservation of energy, we can calculate this quantity. The result is:

$$J_s = J_u/T - \mu_1 J_{m1}/T - \mu_2 J_{m2}/T - \ldots$$

With suitable adaptations this expression can be used to deal with all systems treated so far in the literature. In particular, it can be specialized to the arrangement shown in Figure 3, in which two cylinders are connected by a plug of some sort, perhaps a porous plug, membrane or capillary. This is the simplest system having two terms in the rate of entropy production – a consideration which must outweigh the regrettable lack of suitable experimental data about such systems.

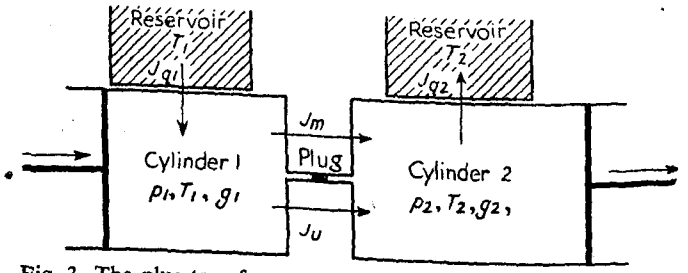


Fig. 3. The plug-transfer experiment: transfer of mass and energy between two cylinders. (The mean heat flow is $J_q = \frac{1}{2}(J_{q1} + J_{q2})$.)

We assume that only a single constituent is involved and we wish to find the rate of *entropy* production in the plug under steady conditions. We may then write:

(rate of production of mass) = (mass flow out) – (mass flow in)

(rate of production of energy) = (energy flow out) – (energy flow in)

(rate of production of entropy) = (entropy flow out) – (entropy flow in).

The conservation laws for mass and energy tell us that the rate of production of these quantities is zero, so that from the first two equations mass and energy flows are equal on either side of the plug. Let us call them J_m and J_u, respectively. Using the expression for entropy flow which has just been derived – and remembering that for a single constituent $\mu = g = u + pv - Ts$ – we find from the equation for entropy balance that:

$$\text{rate of production of entropy} = J_u \Delta(1/T) - J_m \Delta(g/T)$$

This expression confirms the statement made earlier about its form: it is, namely, the sum of two terms each of the type (flow) × (force).

We now have to introduce the assumption (corresponding to the law of heat conduction) that there are linear relations between the flows and the forces. Suppose that they are

$$J_u = A\Delta(1/T) + B\Delta(g/T)$$
$$J_m = C\Delta(1/T) + D\Delta(g/T)$$

where A, B, C and D are constants called 'flow coefficients'. Onsager's principle is that coefficients which are mirror images of one another in the 'leading diagonal' of the right-hand side are equal. In our case, therefore, the principle implies that $B = C$ so that the number of *independent* coefficients is reduced from four to three.

To see more clearly which coefficients are to be equated, it is convenient to introduce temporarily a slight change in notation. Let the entropy production be written in the form $J_1X_1 + J_2X_2$ where J_1, J_2 are any 'flows' and X_1, X_2 the corresponding 'forces'. The assumed linear law – valid near equilibrium – can then be written in these terms as

$$J_1 = a_{1,1}X_1 + a_{1,2}X_2$$
$$J_2 = a_{2,1}X_1 + a_{2,2}X_2$$

The Onsager principle now requires that $a_{1,2} = a_{2,1}$ or (in words): the *contribution to the flow* J_1 due to a unit force X_2 is equal to the contribution to the flow J_2 due to a unit force X_1. The reason why the principle is sometimes called the 'reciprocity

theorem' is now obvious – as is the extension to more than two pairs of forces and flows.

APPLICATIONS OF RECIPROCITY

At first sight the Onsager principle may appear rather esoteric and unpractical. This it has not proved to be. But in seeking examples of its use – and verification – there is a difficulty of another kind. On the one hand, it is possible for plug-transfer experiments to arrive without too much complication at equations relating quantities which are physically measurable, and which in principle can be verified.* But if we ask whether the reciprocity relation has been verified under practical conditions, then we find that there are no very satisfying figures from this type of experiment. If, on the other hand, we go instead for applications in which the experimental verification has been satisfying, then the theoretical derivations of the relations verified is found to be complicated. It may be best, therefore, to give a largely superficial account in both cases.

Analysis of the plug-transfer experiment, illustrated in Figure 3, leads us to expect a rather striking effect which has been observed under quite ordinary conditions. Suppose that the volumes of both cylinders are fixed, but not their pressures, and impose a temperature difference ΔT. Inspection of Figure 3 shows that the system must eventually build up a sufficient head of pressure to reduce the mass-flow to zero. The pressure-head is proportional numerically to ΔT. The constant of proportionality depends both on the fluid used and on the plug, and in many conditions is zero. No one, for example, would expect a

*By changing variables the linear relations given above can be replaced by

$$J_q = -a.\Delta T/T^2 - bv.\Delta p/T$$
$$J_m = -c.\Delta T/T^2 - dv.\Delta p/T$$

where J_q is the mean heat flow (Figure 3). The Onsager principle demands that $b=c$. If this is used with the pair of equations just written then the thermo-mechanical effect (see below) can be related to another experiment by the formula $\Delta p = -(Q^*/vT).\Delta T$, in which Q^* is the mean heat flow per unit mass under isothermal conditions.

pressure head to appear when heat is passed through an ordinary bulk sample of fluid. A positive effect, when found, is known as a 'thermomechanical effect'.

Many attempts have been made to produce a thermomechanical effect by using 'ordinary' fluids, such as water, nitrogen, and methane, and membranes like cellophane and pig's bladder. Although positive results have been claimed, it is only recently that some very skilful work by K. G. Denbigh and G. Raumann in the I.C.I. research laboratories has placed the matter beyond dispute. For their 'plug' they used a latex rubber membrane about 3/10 millimetre thick. They got definite effects from carbon dioxide, nitrogen, and hydrogen. The first-named gave the largest effect: a temperature difference of 12° C. gave a pressure head of 3 centimetres of mercury. It may be thought surprising that such a large effect should have remained undetected for so long. The explanation is no doubt connected with both the difficulty of the experiments and the lack of any previous theoretical incentive to do them.

A second example of a thermomechanical effect has been found in low-temperature physics. If the fluid is liquid helium at a temperature below 2.19° absolute (helium II), and the plug is a capillary tube of diameter less than about 1/100 millimetre, then a thermomechanical effect is obtained. This is the 'fountain effect', one of the famous peculiarities of liquid helium II which are still unexplained in detail.* But there is doubt in this case whether the Onsager theory can be properly applied. We must pass on, therefore, to cases in which the Onsager principle, though technically complicated in its application, has been more adequately tested.

As has already been indicated, Onsager's principle is applicable whenever there are two or more processes occurring simultaneously in a single system. The scope is obviously very wide, and I shall mention the main applications to date under four headings:

*For further information, see Dr Mendelssohn's article, 'Physics of Low Temperatures', *Science News 10*, 71 (1949).

Chemical reactions

When two or more chemical reactions take place together, there is an interaction between them provided that they have reactants in common. In this case the whole import of the Onsager theorem is included in a principle which had been known and used by chemists many years before the reciprocity result was formulated. It is known as the 'principle of microscopic reversibility'* and as a matter of fact was one of the main clues that led Onsager to his goal. In spite of this, the theory of chemical reactions in open systems has benefited indirectly from the spate of research which followed Onsager's discovery.

Conductivity in crystals

Amorphous materials (plastics, liquids, glasses) have the property that if a certain temperature gradient is applied in various directions then the resultant heat-flow will always be of the same magnitude and directed along the gradient. This is not true for single crystals. If you apply a temperature (or electric field) in any direction (x) in a crystal, then in general you will find a component of heat flow (or electric current) in a perpendicular direction (y). It is a notable fact – known for at least a century – that if you now apply the *same* gradient in the y direction then the component of flow in the x direction will be just the same as it was originally in the y direction. The reason for this curious reciprocity – which is accepted without thought by most people – is far from obvious. It can, however, be accounted for by the Onsager principle, and in this case (as for chemical reactions) the experimental evidence is impressive and favourable.

Transport phenomena

Under ordinary conditions we associate a flow of heat with a temperature gradient. But we saw that during transfer between

*We must emphasize in this context that the expression used here has no direct connexion with reversible chemical reactions. For further information the reader may consult Denbigh's monograph, listed under *Further Reading*.

two cylinders a temperature gradient might give rise to a flow of matter, which is normally caused by a pressure gradient. Such a circumstance is referred to as 'coupling' between the heat flow and the mass flow. If we write down the four main 'flows' and their associated 'forces':

flow of		*gradient of*
heat		temperature
matter	associated	pressure
electricity	with a	electrical potential
matter		concentration (chemical
(diffusion)		potential)

we can say that under suitable conditions *any* of the forces can induce *any* of the flows. In the case of transfer between two cylinders we were concerned with the first two pairs of this list and our argument consisted (in essence) in relating the heat flow under a pressure gradient to the matter flow under a temperature gradient. We see immediately that there are at least six such relations possible. Actually there are more, because the diffusion flow recorded above conceals a number of special cases. Also, the theory can deal with more than two pairs at a time.

One of the most interesting cases is the combination of heat flow and electric current. The result of the Onsager law turns out to be a famous relation found by Lord Kelvin in 1900 between the electromotive force of a thermo-couple and the heat given out or absorbed at the junction of two metals. Now it has long been known (and indeed Kelvin himself knew it) that his 'proof' was unsound. It had therefore seemed remarkable that not only was the result correct, but that similar unsatisfactory methods were used with some success to deal with other pairs on the above list. The Onsager principle has provided a unified account of those older results which are correct, and has also predicted a few new relations.

An example of a new relation is found in the combination of diffusion and heat flow. If a fluid mixture is placed in a container in which a pair of opposite walls are held at different temperatures, then when conditions are steady the concentration is found to vary across the container; one of the components

will tend to move towards higher or lower temperatures as the case may be. This phenomenon (sometimes called the 'Soret effect') was tried out as a method of separating U_{235} from natural uranium in making the first atomic bombs. The Onsager principle allows us to express the change in concentration in terms of the energy carried by the diffusion current. The ideas here are clearly very closely related to those involved in plug transfer.

Finally, we may mention that the change in concentration at the boundary between two electrolytes gives rise to a difference in electrical potential – called the 'diffusion potential' – which is of importance in electrochemistry. Reciprocity enables us to find this potential in terms of the mobilities of the ions present.

Electricity in metals

Most textbooks on electricity include a list of curious effects with exotic names: the Nernst effect, the Ettingshausen effect, the Righi-Leduc effect, etc. They all arise from coupling between the magnetic field, the electric field, and the thermal gradient. P. W. Bridgman had found one relation between them, and H. B. Callen, using the Onsager principle, has given it a sound basis and discovered another relation.

As can be seen from this list, the reciprocity law has succeeded in unifying the treatment of a wide variety of effects. Even if the experimental verification is not as broad as it should be, there can be little doubt that the principle is essentially correct. It must be remembered, however, that apart from the concrete predictions indicated above, discussions around the law have led to a deeper understanding of irreversibility. They have changed our way of thinking. Such changes may be less definite but are no less real than mathematical formulae. The consequences of this particular change are not yet exhausted.

SOME GENERAL ASPECTS OF IRREVERSIBILITY

If we attempt to explain the main features of the change, we may expect to see at the same time why irreversible thermodynamics is 'not quite respectable'. The characteristic restriction on classical thermodynamics is that all changes that can be treated quantitatively are reversible changes between equilibrium states. It is, however, true that irreversible changes play an essential role in the classical theory: this is immediately evident from the statement of the Second Law given earlier, and as a matter of fact it can be shown that without irreversibility it would be impossible to define an *absolute* temperature scale. Nevertheless, the introduction of temperature (and, of course, of entropy and the other thermodynamic variables) depends on the concept of equilibrium. What, then, do we mean by talking about the temperature in a non-equilibrium situation?

For 200 years before classical thermodynamics reached its final logical shape, the idea of temperature was firmly established. But if you had asked Fahrenheit what he meant by temperature, it is unlikely that he would have used the word 'equilibrium' in formulating his reply. He would probably have said something to the effect that it was the number indicated on his alcohol thermometer. Our first answer to the above question is likely to be: 'Forget about logical niceties and join Fahrenheit. Temperature is what is measured on a suitable instrument.' It is clear that for such a definition to be reasonable the instrument must indicate the correct temperature for a system in equilibrium. The trouble is that, while a number of different types of thermometer can be found which will do this, they will in general disagree with one another when placed in a non-equilibrium system.

The root cause of this sort of unsatisfactory behaviour is found in the fact that different types of microscopic processes may affect the measuring element in different ways. When we talk of the temperature of the air we normally think of a detecting element that is sensitive to the collisions of the air molecules. But a blackened thermometer bulb placed in sunlight will give a much higher reading than a plain one in the shade. This is because the blackened bulb absorbs more radiation.

It is thus clear that there is no *general* way of defining temperature in a non-equilibrium situation. We can make some progress by restricting ourselves to one type of molecular process. That of greatest interest is the random thermal motion of molecules. Although difficulties of principle remain, the practical

difficulties are much attentuated, because it is precisely this type of process which 'ordinary' thermometers detect. Indeed, we all measure and talk about the temperature of flowing materials (including air) without giving the matter a second thought. For those who like a more formal definition of this commonsense use of the word 'temperature', perhaps the best that can be done is the following: to find the temperature at a point in a moving fluid medium, isolate suddenly a small element around the point and allow it to come to equilibrium; the temperature then measured is defined as the temperature at the point.

The discussion of pressure must follow very similar lines to that of temperature. In this case there are rather fewer methods of carrying out the measurement. Those that do exist are less direct, and similar ambiguities persist in the end. Again we are driven to explain discrepancies in terms of kinetic theory, and again the practical difficulties are less acute than the difficulties of principle.

The case of entropy is different. Entropy is not measured directly nor is it easily 'intuited'. It is rather defined in terms of other quantities – temperature, pressure, and so forth. The assumption made in using the Gibbs equation is that the entropy always depends on these quantities in the same way whether or not the system is in equilibrium. This dependence is known for the equilibrium case (by classical methods) and hence can be written down for the non-equilibrium case. How do we know that the quantity we arrive at *is* the entropy? As it stands, this is not a fair question but is similar to the lawyer's trick: have you stopped beating your wife yet?

Classically, entropy is defined only for equilibrium states. We are therefore at liberty to choose *any* definition during a change of state as long as it reduces to the correct equilibrium answer. For example, we could add to the entropy as defined by the Gibbs equation a function of the temperature gradient which is zero in equilibrium. The resulting quantity could be taken as the entropy without altering any classical results. If we did this, we would destroy the 'thing-like' character of entropy mentioned earlier. In particular, we might find that the entropy of an

isolated system could decrease during its transition from one equilibrium state to another – although, of course, the final state would have a greater entropy than the initial one.

The test of our definition of entropy must lie in its usefulness as an aid to thinking. This is certainly satisfied by the 'equilibrium' definition that we have used. A more stringent test can be applied to gases. For these simplest and best understood of all substances, Boltzmann gave a definition of entropy in terms of their molecular structure. This definition is applicable to both equilibrium and non-equilibrium situations. It is therefore satisfactory that I. Prigogine has been able to show that it agrees with the macroscopic definition as long as the gas is not too far removed from equilibrium. He has even been able to say what is meant by 'not too far' in this context.

Some progress has been made towards interpreting the Onsager theorem in molecular terms. For transport processes which have already been explained in some detail (gas theory, electrons in metals) it appears automatically as the result of the reversibility of molecular motions. The same idea has been used to attempt a more general explanation, but there remains some disagreement over the details.

Philosophically minded readers will have noticed a curious inversion in the above discussion. The early history of our knowledge of matter was naturally dominated by macroscopic concepts – witness the comparatively rapid perfecting of thermodynamics. With the growth of atomic theory the game became to explain these well-understood notions in terms of molecules. This was played so successfully that when (as above) we are worried about lack of clarity in macroscopic concepts, our immediate impulse is to to rush to the microscopic world for aid. While the ultimate aim must be to understand everything – big and small – it may be that we have moved too fast. The microscopic theory of flow processes seems to have reached a plateau, so it is not unreasonable to suggest that at the present time a greater emphasis on phenomenology is likely to prove rewarding.

FURTHER READING

Classical thermodynamics

There are surprisingly few readable textbooks on thermodynamics. One of the best is:

Zemansky, M.W. *Heat and Thermodynamics,* 3rd ed. New York, McGraw-Hill, 1951.

A clear and concise account of the axiomatic foundations of thermodynamics will be found in:

Landé, A. *Handbuch der Physik,* **9,** 281 (1926).

Onsager reciprocity

An excellent, simplified treatment of the leading ideas with particular emphasis on their physical content is given in:

Denbigh, K. G. *The Thermodynamics of the Steady State.* London, Methuen, 1951.

For a more comprehensive and mathematical account, with full references, by the man who has, perhaps, done more to elaborate the theory than anyone else, *see*:

De Groot, S. R. *Thermodynamics of Irreversible Processes.* Amsterdam, North-Holland Publishing Company, 1951.

A brief but cogent review, which emphasizes the importance of stationary (time-independent) states, is:

Haase, R. *Zeitscrift für Naturforschung,* **6a,** 420, 522 (1951).

The thermomechanical effect

For important experiments cited in the text *see*:

Denbigh, K. G., and Raumann, G. *Proceedings of the Royal Society,* **A 210,** 377 and 518 (1951).

Three papers on liquid helium which give some idea of both the difficulty and the fascination of experiments in this field are:

Kapitza, P. L. *Journal of Physics, USSR,* **4,** 5 (1941).

Mellink, J. H. *Physica,* **13,** 180 and (with Meyer, L.) 197 (1947).

Thermodynamic description of non-equilibrium states

A stimulating non-technical essay on such topics as entropy-as-a-thing and the definition of temperature, with some amusing paradoxes (unresolved), is:

Bridgman, P. W. *The Nature of Thermodynamics.* Boston, Harvard University Press, 1942.

For the calculation of entropy production and applications to the efficiency of engines, and for a discussion of temperature measurement away from equilibrium, see the following reference (with the warning that the treatment of thermo-electricity given here is fallacious):

Tolman, R. C., and Fine, P. C. *Reviews of Modern Physics,* **20,** 51 (1948).

HISTORIC PHOTOGRAPHS

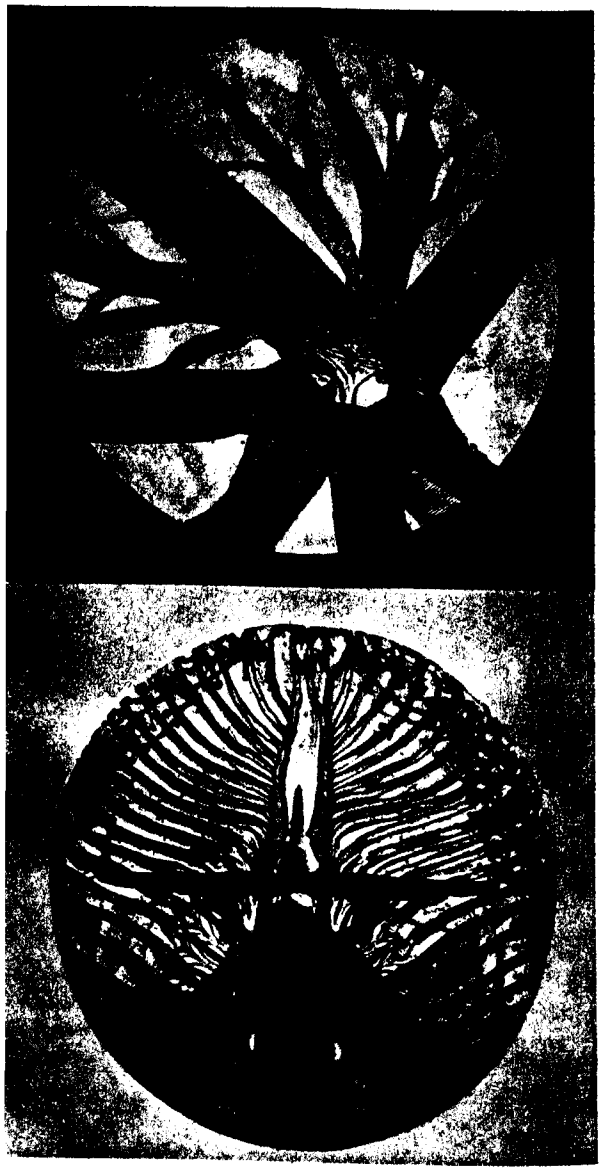
THE centenaries of photography and, this year, of the Royal Photographic Society have been celebrated in a number of exhibitions which have illustrated the progress of photography both technically and as an art. There seems to have been no exhibition of photographs which are historic in relation to the many branches of science in which photography has been used as a tool. A number of such photographs are reproduced as Inset 1–20. Many more could have been included; there are, indeed, whole areas of scientific photography which are not represented. Most of the photographs are connected with specific discoveries; others are included as the first or early examples of new techniques. The period covered is 1845 to 1952.

It is convenient in relation to dates to recall the three main stages in the development of photography as a widely practised technique. During each of these stages, the primary process in the taking of a photograph has been the reduction of silver salts (iodide and bromide) to metallic silver on exposure to light; and the most important changes in the use made of this process have been changes in the backing and/or medium supporting the light-sensitive salts. For the working scientist, these changes have been as significant in relation to ease of operation and reproduction as in relation to quality and sensitivity; only at the beginning of the last stage was a sudden advance in sensitivity the result of a new type of plate. The beginning dates of the three stages are taken usually as 1839, 1851, and 1880.

The first stage, from 1839 to 1851, can be described broadly* as that of pre-glass-plate photography. Two different methods

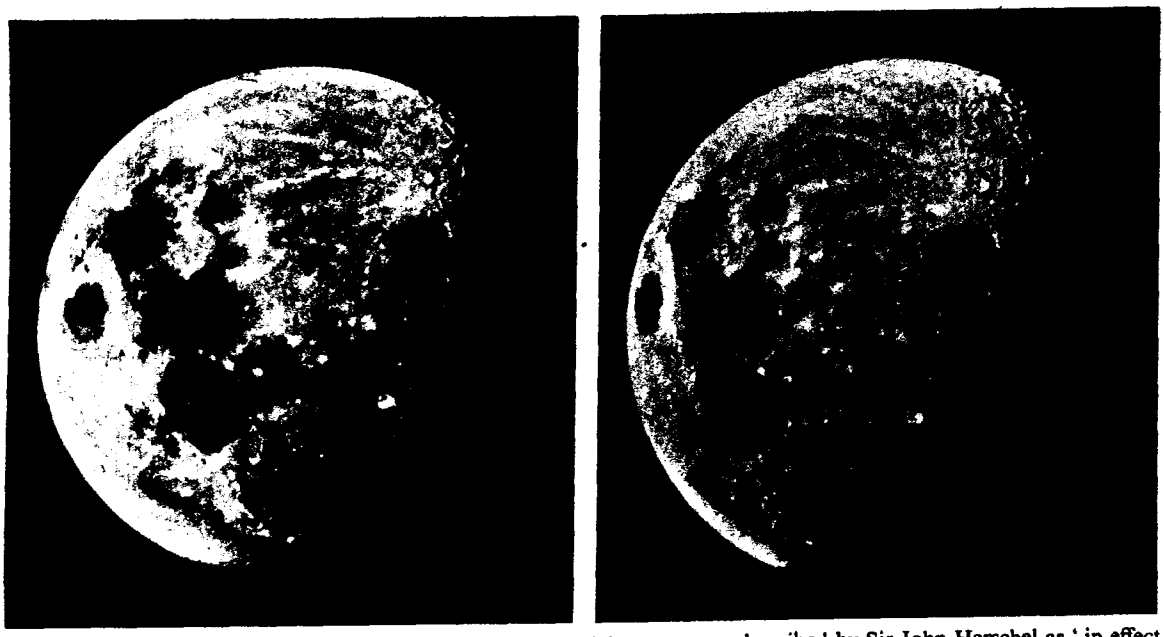
*The qualification is necessary because of a negative on glass obtained by Sir John Herschel in 1839. It showed the 40-foot telescope built by his father at Slough. The plate was prepared by the all but impossible process of precipitating silver chloride from solution on a glass plate and syphoning off the liquid with as little disturbance as possible.

1. Daguerreotype-micrograph of the blood corpuscles of the frog. It was made by Léon Foucault for the *Atlas* of Alfred Donné, published in 1845. Magnification × 500. (*Courtesy of The Wellcome Historical Medical Museum.*)

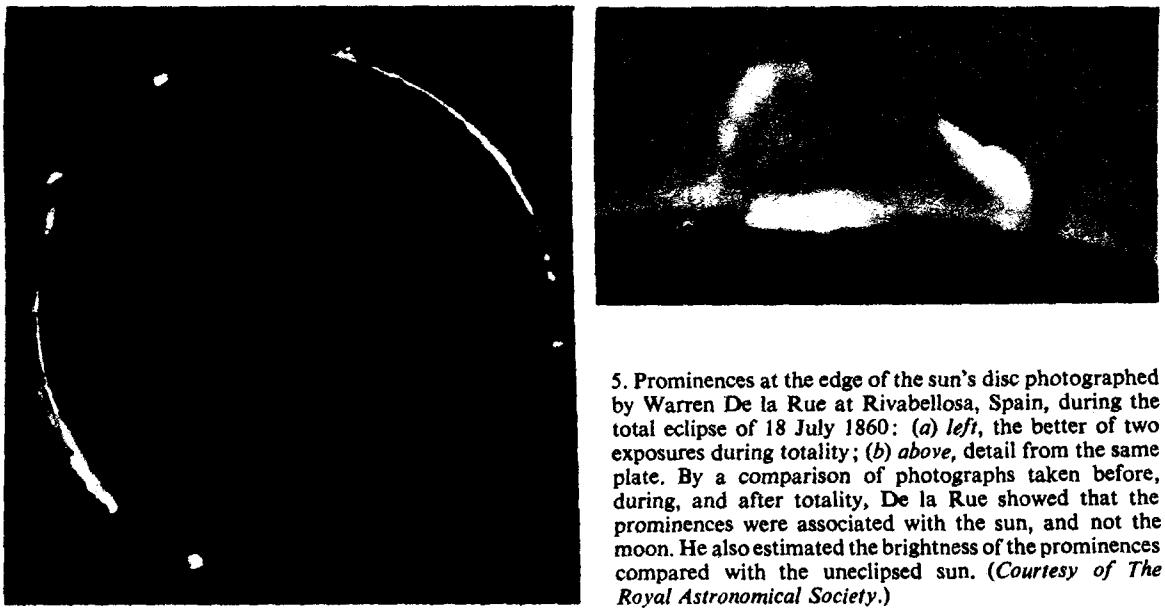


2 (*left page*). The first photomicrographs by the wet collodion process: (*a*) *above*, spiracle and tracheae of the silkworm; magnification × 55. (*b*) *below*, proboscis of the fly; magnification × 160. Joseph Delves of Tunbridge Wells; published as single plate in the Transactions of the Microscopical Society, 1853. See also above. (*Gernsheim Collection.*)

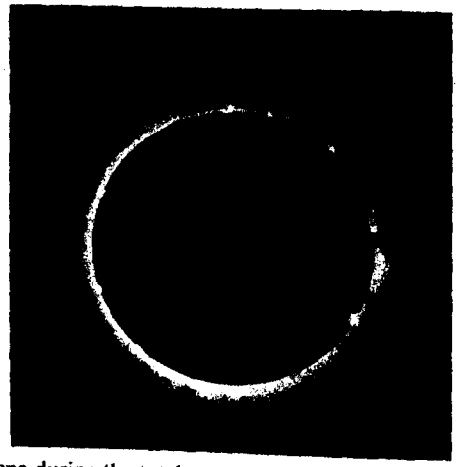
3. Proboscis of the fly; detail from Inset 2 (*b*). Magnification × 500.



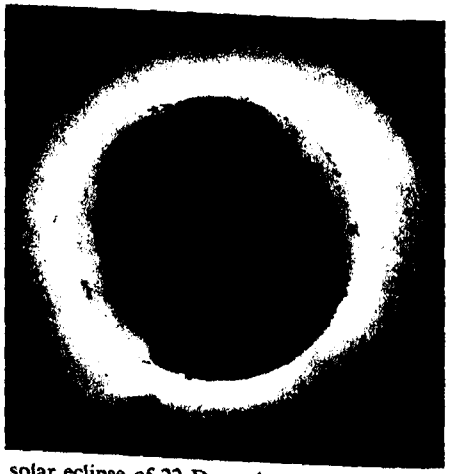
4. The extension of the stereoscopic principle to the surface of the moon was described by Sir John Herschel as ' in effect a step out of and beyond nature ': stereoscopic pair by Warren De la Rue, 1857. (*Gernsheim Collection.*)



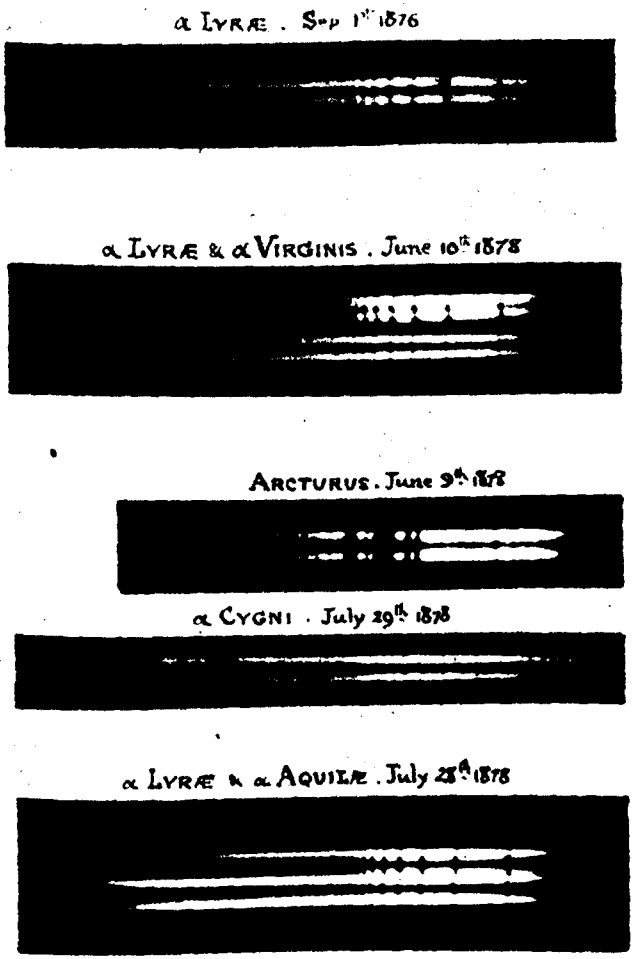
5. Prominences at the edge of the sun's disc photographed by Warren De la Rue at Rivabellosa, Spain, during the total eclipse of 18 July 1860: (*a*) *left*, the better of two exposures during totality; (*b*) *above*, detail from the same plate. By a comparison of photographs taken before, during, and after totality, De la Rue showed that the prominences were associated with the sun, and not the moon. He also estimated the brightness of the prominences compared with the uneclipsed sun. (*Courtesy of The Royal Astronomical Society.*)



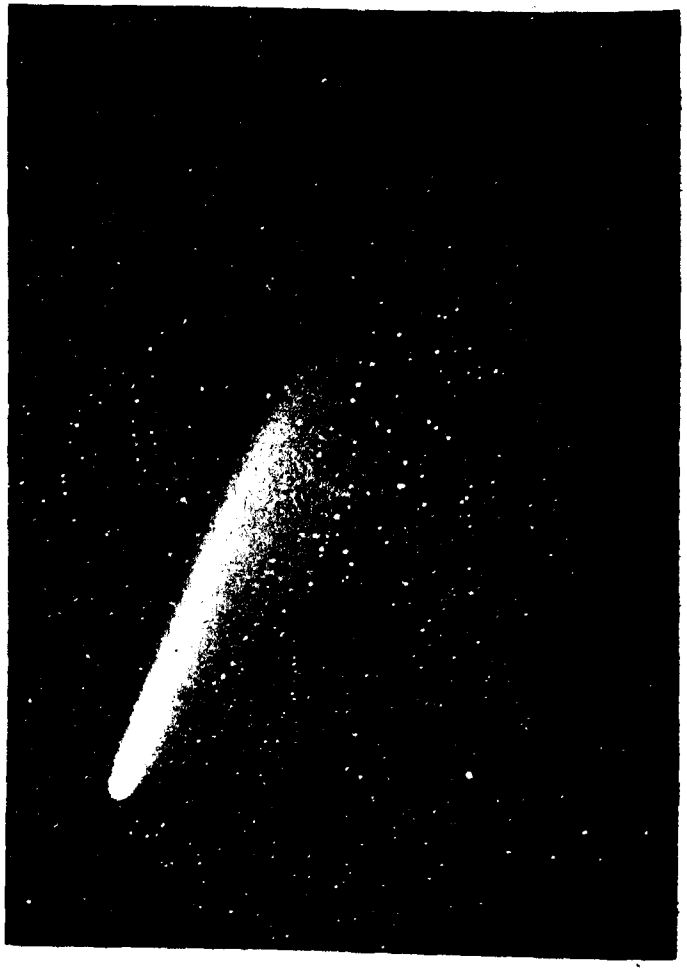
6. The corona during the total solar eclipse of 28 July 1851: the reproduction is from a woodcut based on a Daguerreotype which was taken by Dr. A. L. Busch at Königsberg and provided the best evidence of the form of the corona. (*Courtesy of The Royal Astronomical Society.*)



7. The total solar eclipse of 22 December 1870: photograph taken at Cadiz by Professor C. A. Young of Princeton University. Reproduction is from a negative copy on which the positions of prominences have been lightly inked, probably for comparison with other photographs and drawings. (*Gernsheim Collection.*)

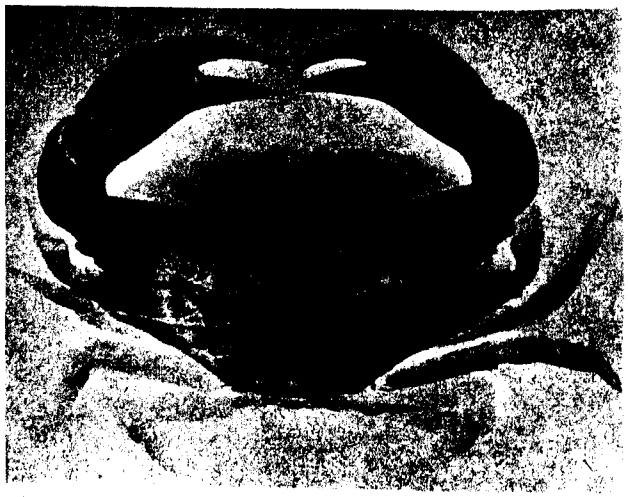


8. Stellar spectra photographed by Sir William Huggins between 187 and 1878, and included as 'historic' in the *Atlas of Representati Stellar Spectra* which he published jointly with Lady Huggins in 189

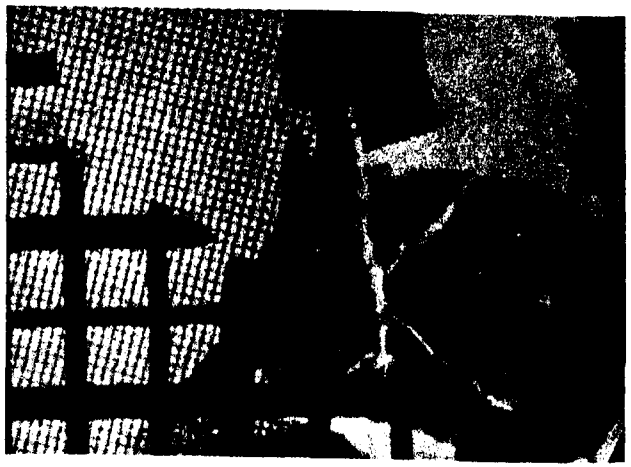


9. The 1882 comet photographed by David Gill at the Cape Observatory, with a photographer's camera strapped to an equatorial telescope. The number of stars shown on this and other plates contributed to the international decision, taken in Paris in 1887, to undertake the *Carte du Ciel*. (*Courtesy of The Royal Astronomical Society.*)

10. Radiograph of right hand, with a piece of a needle lying close to th radial side of the second phalanx of the index finger; taken by Sydne Rowland between 4 March and 6 March 1896. This is one of earlies radiographs taken for medical use which is also satisfactory as a radio graph. (*Courtesy of The British Institute of Radiology.*)

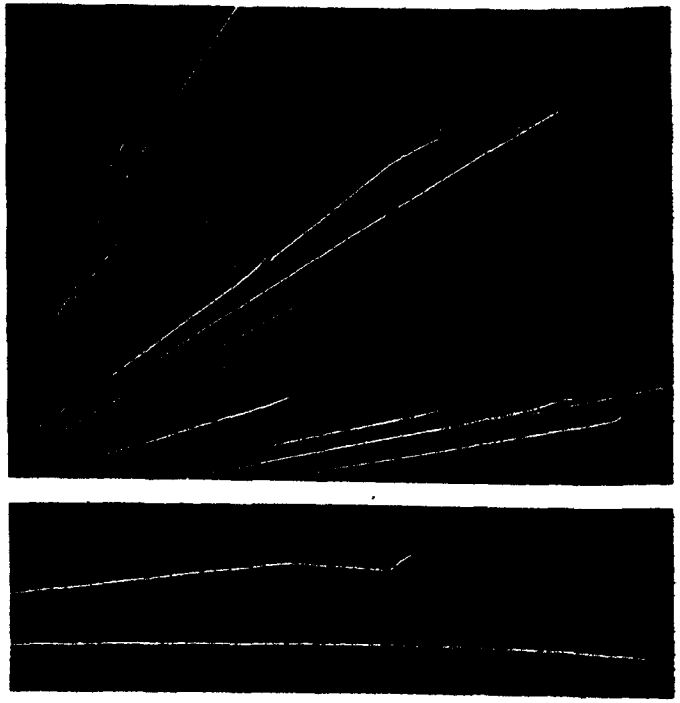


. Radiograph of crab. By R. Norris Wolfenden, February 1897. *'ourtesy of The British Institute of Radiology.*)

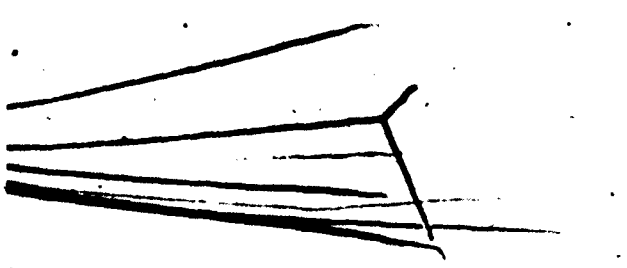


. Experimental X-ray micrograph, one of the first with biological ıterial included: silver grid of 1500 lines per inch, and cut copper d, over frog kidney section, 0·1 millimetre thickness, stained with osphotungstic acid. Magnification × about 75. (*V. E. Cosslett and . C. Nixon. April 1951.*)

13. X-ray micrograph of Drosophila melanogaster. Magnification × about 55. Exposure, 5 minutes ; the specimen had been freeze-dried. (*Cosslett and Nixon. Courtesy of the Royal Society.*)



14. Cloud-chamber photograph, published by C. T. R. Wilson in 1912. Tracks of alpha-particles from radium: (*a*) *above*, showing that most of the tracks are straight; (*b*) *below*, enlargement of part of the photograph to show a track with two abrupt changes of direction. The comparative rarity of such collisions provided supporting evidence for the Rutherford theory of the atom.

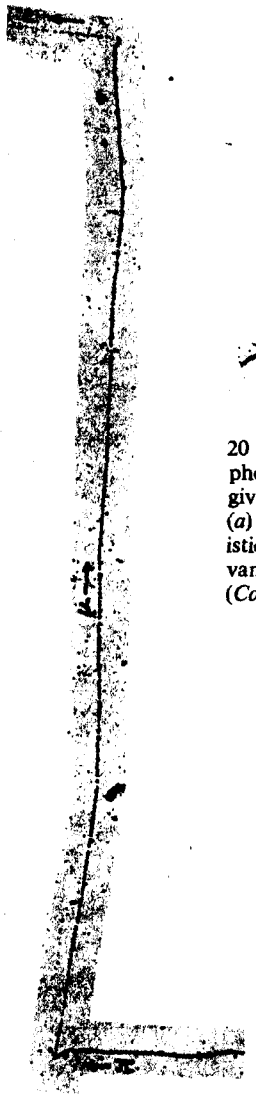


15. Cloud-chamber photograph: one of the 8 examples of forked tracks found by P. M. S. Blackett in 1925 from more than 20,000 photographs of alpha-particle tracks. (See text.)

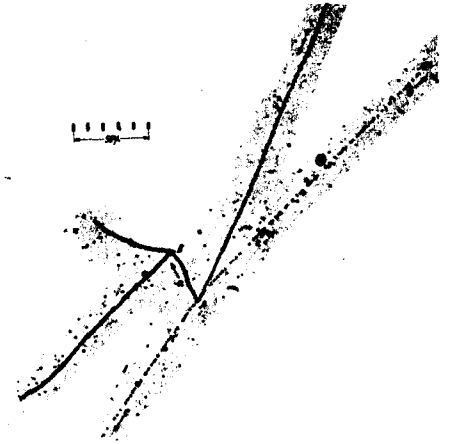
16. Artificial disintegrations of lithium and boron atoms produced by J. D. Cockcroft and E. T. S. Walton in 1932 were photographed by P. I. Dee and Walton by the cloud-chamber method. The picture shows the transmutation of boron into three alpha-particles, indicated by arrows. (*The photographs on these two pages are reproduced by courtesy of the Royal Society.*)

17 (*above*). Cloud-chamber photograph of tracks of positive and negative electrons; it was obtained by P. M. S. Blackett and G. P. S. Occhialini in 1932 by a newly developed method in which high-energy particles were made to take their own cloud photographs. The positive tracks are those with curvature to the right. (*Courtesy of the Royal Society.*)

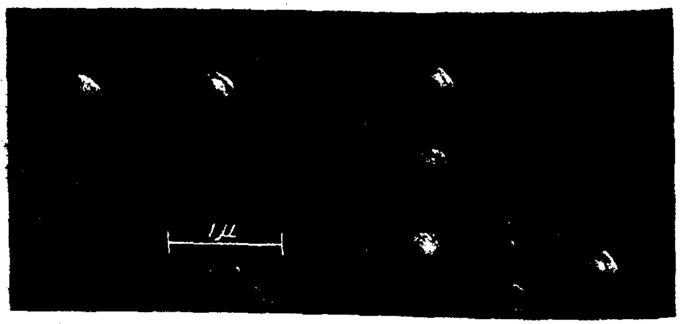
18 (*right*). Nuclear emulsion method: photomicrograph mosaic of the first recorded example of the breakdown of a positive π-meson into a μ-meson. It was obtained by C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini and C. F. Powell in 1947. (*Courtesy of the Editors of* Nature.)



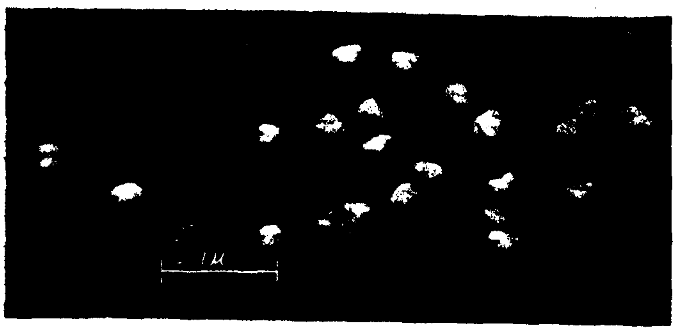
19 (*left*). Nuclear emulsion method: with the development of electron-sensitive emulsions, the breakdown of the positive π-meson could be followed a stage further; this is the first photomicrograph showing the successive breakdowns, positive π-meson to μ-meson, μ-meson to positive electron. (*Courtesy of Professor C. F. Powell, F.R.S.*)



20 (*above*). Nuclear emulsion method: first photomicrograph showing a τ-meson. This gives rise to three particles, one of which (*a*) produces a nuclear disintegration characteristic of a negative π-meson. Observer, Mrs van der Merwe; R. Brown and others, 1949. (*Courtesy of the Editors of* Nature.)



21. Electron micrograph of purified myxoma virus, ' shadow-cast ' with uranium. Magnification × about 15,000. (*Photograph by courtesy of J. Farrant, Chemical Physics Section, Division of Industrial Chemistry, Commonwealth Scientific and Industrial Research Organization.*)



22. Electron micrograph of vaccinia virus, taken under the same conditions and reproduced at the same magnification. (*Farrant and Fenner.*)

were available: photography on metal, the 'Daguerreotype', in which the initial material was a silvered copper plate; and photography on sensitized paper, the method usually associated with the name of W. H. Fox Talbot. The year 1839 saw the publication by Louis Daguerre of the daguerreotype process, and by Fox Talbot of 'photogenic drawing', which gave negative pictures, followed a year later by the 'Calotype' process, in which 'positives' were obtained by contact-printing from paper negatives. To the extent that photography was used in scientific work during the next twelve years, the daguerreotype was mostly used. It offered a brilliantly defined image, but the plates of silvered copper were relatively expensive, processing was complicated and difficult, and only one picture was obtained. Exposure-times, which at the beginning were around half an hour, were later reduced to only a few seconds under favourable conditions, and there was no gain in sensitivity from the transition to the next method. For the 'reproduction' of a daguerreotype, it was usual to draw from the original and engrave.

The second stage, beginning in 1851, followed quickly from the introduction of the 'wet collodion' process by F. Scott Archer. In this the backing material was glass; the light-sensitive salts were formed in collodion, a solution of nitrocellulose in alcohol-ether; and the plates were prepared immediately before use, and were exposed wet. The method offered enough improvement, and particularly the facility of reproduction, to oust the daguerreotype (and also the calotype) within a few years. Where the advantages of using photography in research were obvious and immediate, the new process was quickly adopted. But for incidental use there was the discouragement that the plates had to be prepared in a darkroom immediately before exposure. And there was a restriction on exposure-time from the requirement of exposure while still wet.

The third date, 1880, is that of the advent, in general use, of dry plates of essentially modern type, following experiments by W. L. Maddox, in 1871, on the making of gelatine emulsions. Compared with the rapid sweep effected by the wet collodion process, this was relatively a slow transition; but, by 1875,

gelatine-emulsion plates were in use in research, and, by 1879, satisfactory plates could be bought commercially. Full mobility had been given to the camera; and, in most uses in research, it was no longer necessary that the working scientist should be also a photographic chemist. The only remaining development of general importance was the extension of the range of wavelengths to which plates would respond, and even this problem had been solved in principle by 1873.

EARLY SCIENTIFIC USES

The earliest scientific uses of photography were in microscopy and astronomy; that in microscopy being the first by some two years.

The first photomicrographs were taken by the Rev. J. B. Reade in 1837. He used a solar (projection) microscope, and his pictures were negatives, an independent invention of the 'photogenic drawing' of Fox Talbot. Some of them were shown at a Conversazione held by the Royal Society in London in 1839. He showed a number of biological subjects, one of them a flea, at magnifications which ranged from 50 to 150. Novel though they were, they must have possessed some merit as pictures to have aroused this amount of attention.

One of the first serious uses of the daguerreotype was also in microphotography. Inset 1 is a photograph of one of a series of daguerreotypes made by Léon Foucault in Paris for the *Atlas* of Alfred Donné, published in 1845; the *Atlas* being itself a supplement to Donné's *Cours de Microscopie,* published in the previous year.

From its modern reproduction the daguerreotype might be thought to have been almost as useful to the research worker as its successor, the wet collodion process. The first two photomicrographs produced by the latter method are shown as Inset 2; and an enlargement of part of the lower micrograph, the proboscis of the fly, as Inset 3. The negatives were made by Joseph Delves at Tunbridge Wells in 1852, only one year after the introduction of the process. The 'positive proofs' were made by Mr Henneman of Regent Street, and the present reproductions are

from a photograph of one of his century-old prints. The two micrographs were first published, with only slight loss in quality, in the *Transactions of the Microscopical Society* for 1853. As well as a short note by Mr Delves, there was a 16-page account by Samuel Highley jun., 'as one of the publishers', of the principles, processes, and apparatus which were involved in the first practical application 'as a printing process to the illustration of scientific literature'.

With a few exceptions in astronomy, daguerreotypes made of physical subjects were important rather for the interest which they created than for intrinsic value. Daguerre himself made a picture of the moon as early as January 1839; he did so at the suggestion of Jean Baptiste Biot, at that time one of the most influential of French physicists. He was followed in 1840 by Dr J. W. Draper of New York; and, in 1842, Draper recorded the first diffraction spectrum. The latter is now held in a country store by the Science Museum, London. Foucault, fresh from his work for Donné's *Atlas,* joined with Fizeau in 1845 to obtain a picture of the sun. Later daguerreotypes of the moon were made by W. C. Bond and J. A. Whipple at the Harvard Observatory, and were exhibited in London in 1850, and again in 1851 at the Great Exhibition. They were seen on the latter occasion by Warren De la Rue; and thus, in England as at Harvard, paved the way for 'celestial photography' using the wet collodion process.

A daguerreotype which was of some importance in its own right was one of the sun's corona taken by A. L. Busch at Königsberg during the total eclipse of 28 July 1851. This was the first representation of the corona which was thought to be worth including in a memoir on 'Observations Made During Total Eclipses'* published by the Royal Astronomical Society in 1879. It provided the chief evidence for the form of the corona at this eclipse, and was discussed also as providing the first evidence at any eclipse for the approximate symmetry of the corona with respect to the axis of rotation of the sun. Inset 6 is reproduced from a woodcut which was in turn made from a steel engraving

Memoirs of the Royal Astronomical Society,* **41 (1879).

of this daguerreotype. It is thus a third-hand reproduction, although it was checked against the original before its publication in 1879 and was found to represent it 'very fairly'.

The impact of the wet collodion process on microphotography has been mentioned already. The new process was taken up with equal speed in astronomy. By 1852, the year following its publication, it was being used by Warren De la Rue in England, and by the Harvard astronomers in the United States. Soon after, Father Angelo Secchi in Rome joined these early pioneers.

De la Rue was an amateur, and it was his good fortune that the incentives provided by the Harvard daguerreotype and the new process came to him at the same time. Within a few years he had obtained many excellent photographs of the moon. Interest in stereoscopic viewing was increased a year later by Charles Wheatstone, whose adjustable stereoscope was evolved in 1852. Wheatstone saw that the slight apparent oscillation of the moon's disc known as libration afforded an opportunity for stereoscopic viewing. A stereoscopic pair, obtained by De la Rue, is reproduced from a contemporary transparency as Inset 4. The method was described by Sir John Herschel as a 'crowning triumph . . . in effect a step out of and beyond nature', the moon being exhibited 'as it would be truly seen by a giant whose eyes were 52,000 miles asunder, if stationed at our distance from it'.

For the routine observation of sunspots, De la Rue designed a photo-heliograph which was installed at Kew Observatory in 1858. Within two years he was borrowing the apparatus for his own experiments. He took it first to Rivabellosa in Spain to obtain the historic photographs of solar prominences seen in total eclipse, one of which is reproduced as Inset 5; and in the two following years for work at his private observatory. There he obtained the first stereoscopic presentation of sunspots, showing the spots as holes. Slides made from his plates are still in existence, but have developed markings of their own which would not be readily distinguishable from sunspots.

Inset 5 shows the more successful of two photographs taken during the total phase of the 1860 eclipse. The main value of the expedition was the evidence provided by successive exposures

before, during, and after totality that the prominences were associated with the sun's surface and not the moon's. The telescope was found to have been jolted twice in the course of the other exposure during totality, the effect of three superimposed photographs being produced. The accident was used by De la Rue to make a first estimate of the amount of light emitted by the prominences.

The first wet-collodion photographs showing any substantial part of the corona were taken during the same eclipse with Father Secchi's telescope at Desierto de las Palmas, also in Spain. They supported De la Rue's conclusion about the prominences, but in relation to the eclipse observations as a whole they added little about the corona. Inset 6 shows a photograph taken at Cadiz by Professor C. A. Young of Princeton during the 1870 eclipse. It was at this eclipse that a Manchester astronomer, Alfred Brothers, confirmed photographically that the corona, like the prominences, is associated with the sun and not the moon. The slide from which Inset 6 is reproduced was formerly in the possession of Brothers; a number of identifiable features round the sun have been outlined on it in ink, probably as an aid to comparison with his own photographs.

The use of photography to determine the positions of stars is a curious example of an apparently obvious application, which was suggested in the first stage of enthusiasm, was soon afterwards investigated and found to present difficulties, and finally was revived almost by chance after dry plates became available. The first suggestion was made in 1857 by G. P. Bond of the Harvard Observatory. It was investigated by De la Rue, and was discussed by him at the British Association meeting of 1861. He pointed out that 'the instrument best adapted for this purpose is a camera of short focal length, compared with its aperture, like the ordinary portrait camera'. However, there were difficulties inherent in the wet-collodion process. One, noted by De la Rue, was that the images of the stars were 'no bigger than the specks common to the best collodion'. Another was that exposures could not be long enough with wet plates for the fainter stars to be recorded. De la Rue tried experiments with photographs

taken slightly out of focus, so that stars and specks could be more readily distinguished. It might have been thought, therefore, that the idea would have been in mind when the gelatine-emulsion type of plate became generally available.

In practice, an independent stimulus was needed. It was provided by some photographs taken of the 1882 comet at the Cape Observatory. Hearing that the comet had been photographed with ordinary cameras, David Gill, then H.M. Astronomer at the Cape, called in an outside photographer and strapped his camera to the great equatorial telescope at the observatory. One of his photographs is reproduced as Inset 9. As lately summarized by Dr Merton: 'He was so impressed with the numbers of stars recorded that he initiated a chain of events, involving too the Henry Brothers of Paris, which led to the historic meeting in Paris in April 1887 of the International Astrographic Congress at which the great project of the *Carte du Ciel*, the "Astrographic", the cataloguing and charting of the whole sky, was undertaken with the co-operation of many observatories, but is even yet not completed.'

There was a similar sequence of early experiments, followed by an interval, in stellar spectroscopy; but with the variation in this case that one man, Sir William Huggins, was working at both stages. The story has been told by Sir William and Lady Huggins* in their account of work done at the private observatory established by Sir William in Upper Tulse Hill, London, in 1856. The significance of the date is that only three years earlier Kirchhoff at Heidelberg had shared with Bunsen in the establishment of spectroscopy as a method of chemical analysis, and had himself inferred the presence of known chemical elements in the sun. Huggins, looking around for a new direction of research, decided 'to extend his [Kirchhoff's] novel methods to the other heavenly bodies', and, in co-operation with Dr W. Allen Miller, the observatory was made 'a meeting place where terrestrial chemistry was brought into direct touch with celestial

**Publications of Sir William Huggins's Observatory. Vol. 1: An Atlas of Representative Stellar Spectra* (London, 1899).

chemistry'. Others also were beginning to work on the problem: Rutherfurd and Henry Draper in the United States, Secchi in Rome, Vogel in Berlin. Of his own contribution at this time, Huggins wrote that 'Star-spectra, as distinguished from each other by their main features, were originally described by myself and Dr Miller in 1863, from a direct comparison of the spectra of some fifty representative stars with the spectra of many of the chemical elements. These observations established the important, and, at that time, new fact, that all the stars belong to the same order of bodies as our sun.'

Out of these fifty stars, the spectra of two only, Sirius and Capella, were photographically recorded. Their spectra were photographed in January and March 1863 on wet collodion plates, and were only partly successful. Of the difficulties encountered, staining caused by the draining and drying of the plates during the long exposures necessary would have been the most troublesome to overcome.

Huggins's interest in photographic recording is shown by the early date – 1875 – at which he returned to the attack, using gelatine-emulsion plates some five years before their general adoption by photographers. 'An important advance in the knowledge necessary for a scientific classification of the spectra of the stars', he wrote twenty years later, 'was made during the years 1875–1879 by means of ... successful photographs.' His two earliest spectra were lost; examples of these later ones, described in the 1899 volume as 'historical', are reproduced as Inset 8.

If a rather large share of space has been allotted to astronomy, it is because this is the branch of science in which photography is most widely essential. In its further development, the history of photography in astronomy is barely distinguishable from the history of astronomy as a whole.

SOME MORE RECENT APPLICATIONS

Of the remaining photographs which have been assembled, the majority are related to modern physics, beginning with the discovery of x rays by Röntgen. The discovery itself was photographic, although a fogged plate, however historic, could hardly

be described as a picture. So great a furore resulted from the general publication, early in January 1896, of Röntgen's discovery that the identity of the first useful radiograph will probably never be established; there is a question in any case as to the degree of usefulness which should be counted. Inset 10 is reproduced as *one* of the earliest radiographs taken for medical purposes* which was also creditable as a radiograph, and is available in a form suited to reproduction. It was taken between 4 March and 6 March 1896 by Sydney Rowland. The case to which it relates was described in Volume 1 of *Archives of [Clinical†] Skiagraphy*, of which the earlier pages were published in May 1896 without formal designation as an issue or part. The patient was a woman of 56, who came to the Royal Free Hospital, London, on 4 March, complaining of having run a piece of needle into the index finger of her right hand three days earlier. The radiograph was taken at the surgeon's request, the position of the needle fragment was clearly defined, and two days later it was decided to operate. What is still remarkable is the quality of the radiograph as one obtained within two months from the news of Röntgen's discovery.

Inset 11 is reproduced as an early example of a radiograph of a zoological subject. The crab illustrated was one of a variety of specimens collected by Dr R. Norris Wolfenden, while dredging near Scapa Flow, Orkney, and examined in February 1897 in the laboratory of Dr Macintyre of Glasgow.

The development of *x*-ray shadow microscopy to the point where it is now a usable technique has been recorded in recent

*It was certainly preceded by, for example, an apparently poor radiograph which was taken with home-assembled apparatus by Sir Oliver Lodge, which sufficed to show the position of a bullet in relation to the bones of the wrist and was jointly published by Lodge and the orthopaedic surgeon, Sir Robert Jones, in the *Lancet* of 22 February 1896. Even four weeks before this, radiographs had been obtained in at least three countries which *could* have been used clinically, but without evidence that they were so.

†Although usually quoted by the longer title later adopted, the word *Clinical* is omitted from the title-page.

issues.* Inset 12 shows one of the first *x*-ray micrographs, by V. E. Cosslett and W. C. Nixon in the Cavendish Laboratory, in which biological material was included. At this stage (April 1951), the initial magnification was by about 15, and the limit of useful photographic magnification by about 100; a 5-minute exposure was given. The present limit of useful magnification is by about 400. Whether these micrographs will be regarded by future generations as historic will depend on the further development of the method and the use made of it by biologists. They are in any case of interest as an extension from classical radiography.

The methods of *x*-ray crystallography have been almost wholly photographic from the beginning. They were illustrated by Professor Kathleen Lonsdale in *Science News 18*, and have been omitted from the present collection of photographs. Similar considerations apply to electron microscopy – of which Inset 21–22 are incidental examples in connexion with another article – and also to phase-contrast microscopy, the use of interference methods in optics and for the measurements of surface features on crystals, and to aerial and high-speed photography. Not all of them could have been represented, and it has seemed better to concentrate on a single field, nuclear physics, in which cloud-chamber photographs and, more lately, the use of 'nuclear' photographic plates have provided complementary methods for the recording of charged particles.

Inset 14 (*a*) is one of the first cloud-chamber photographs published by C. T. R. Wilson. It shows a group of tracks formed by alpha-particles diverging from a radium source to the left of, and below, the picture. It was published in 1912 together with a number of other such pictures, of which it might be said that their real value was that, for the first time, they brought atoms to life. The comparative rarity of tracks, such as that of Inset 14 (*b*), which show appreciable deflexions, provided supporting evidence for the Rutherford theory of the atom with the mass concentrated into a small volume, the nucleus.

Inset 15 is a later example, showing the value for interpreta-

Science News 25, 117 (1952); *Science News 26*, 35 (1952).

tion, as well as conviction, of the individually recorded track. It belongs, in Aston's description, to 'the period of quiescence awaiting the development of new weapons' for attacking the nucleus. In more than 20,000 photographs, depicting more than 400,000 tracks of alpha-particles in nitrogen, P. M. S. Blackett found eight examples of a forked type of track, showing that a nitrogen atom had been disintegrated. The results, published in 1925, provided evidence of the mechanism of disintegration: namely that the bombarding nucleus enters intimately into the reaction, with a compound nucleus formed as the first step.

The end of the period of quiescence is marked by Inset 16. It is one of the cloud-chamber photographs, taken in 1932 by P. I. Dee and E. T. S. Walton, of the disintegration of lithium and boron nuclei by high-energy protons, as then lately effected by J. D. (now Sir John) Cockcroft and Walton. The one reproduced shows the ejection of three alpha-particles from a disintegrated boron nucleus; this is visually more effective than the corresponding photographs showing the ejection of two alpha-particles, in opposite directions, from a lithium disintegration.

The remaining cloud-chamber photograph (Inset 17) should, for the record, have been accompanied by another, by C. D. Anderson, showing the first published track of a positive electron. Inset 17 was one of a number of pictures obtained by Blackett and G. P. S. Occhialini with a new apparatus in which cosmic radiation particles were made to take their own cloud photographs. Because of the slight delay necessary for the expansion of the chamber, there is a loss in sharpness with this procedure compared with instantaneous photography. On the other hand, the proportion of useful pictures was much increased, and in this one photograph there are at least two tracks which were shown decisively by their direction of curvature in a magnetic field to be those of positive electrons. It was again Anderson who made the first observations of meson particles, also by a cloud-chamber method.

Inset 18–20 are examples of different first observations made at Bristol University by the nuclear emulsion method, for which Professor C. F. Powell has been primarily responsible. Inset 18

is the simplest, the earliest, and perhaps the most striking. It was published in 1947, and shows the breakdown of a positive π-meson at the end of its range into a μ-meson of smaller mass. The direction of motion of the π-meson, from bottom to top in the reproduction, is shown by the increasing ionization, and therefore grain density, towards the end of the track. It was a coincidence, but effective, that the μ-meson was ejected in almost the opposite direction, from top to bottom in the picture. Inset 19 shows the same process carried a stage further, as became possible after the production of electron-sensitive emulsions. It was reproduced as Inset 29 in *Science News 16*, in illustration of an article by Terence Price on 'Recent Work on Mesons', in which the method was further described. As before, a positive π-meson breaks down into a μ-meson; but this, in turn, breaks down into a positive electron. It is again coincidence that the angles between tracks are here close to right angles.

In Inset 20, the primary particle is a heavier τ-meson. It moves diagonally downwards from right to left, and, at the end of its range, gives rise to three particles, one of which produces a disintegration characteristic of a negative π-meson. An observational link was thus established between the τ-meson and a particle of which the properties were already adequately known.

CONCLUSION

It need only be added to this catalogue that there is a case for a central collection of prints or negatives of photographs, including daguerreotypes, which are of historic interest in science, and that this should be established sooner, rather than later, before too many are lost, damaged, or deteriorate.

Acknowledgement is made to Mr Helmut Gernsheim and to Dr E. Ashworth Underwood for information; and for help in the collection of material to Dr V. E. Cosslett, Dr R. W. B. Pearse, Professor C. F. Powell, F.R.S., Professor J. G. Wilson, The British Institute of Radiology, the Director of The Cavendish Laboratory, The Royal Astronomical Society, and The Wellcome Historical Medical Museum.

A. W. H.

DIURNAL RHYTHMS IN ANIMALS

J. L. CLOUDSLEY-THOMPSON

THE transition from day to night is accompanied by many changes. Physically there is a fall in temperature and a corresponding increase in the relative humidity of the air as well as a reduction in light intensity. There are also marked variations in the timing and extent of these changes according to the locality and habitat. In a continental desert for example, the difference between day and night temperatures may be very considerable, but in the wetter parts of the tropics it is no more than a few degrees. Again, in the upper layers of lakes and oceans the change in light intensity is comparable with that on land, but the temperature remains relatively constant. At the same time there are changes in the behaviour of plants and animals, some of which are obvious and well known while others are more obscure. The owl and badger come out at night when the hawk and squirrel have retired to rest. Even in worm-like animals, those 'enliveners of wet places', we find marked 24-hour rhythms of activity. Such behaviour could be a response to any of a number of physical changes, but there is reason to think that variations in light intensity are the most widely important.

On the other hand the biological advantages to a species of rhythmic behaviour are often determined by secondary factors. For example, a species may become nocturnal as an adjustment to competition from other species, or to escape being eaten. Thus there are many nocturnal forms in tropical forests where competition during the day is particularly severe. As a short-term adjustment, game animals tend to nocturnal habits when much hunted, and rabbits do not come out until nightfall in areas that are often shot over. Conversely, predators must be nocturnal if their prey comes out only at night.

Many desert animals such as centipedes, spiders, scorpions, mites, ants, beetles and other insects, lizards and snakes avoid the extreme mid-day heat, drought, and ultra-violet light by hiding under rocks and stones or in cracks in the ground. They soon die if taken from their burrows and placed on the hot sand. Birds take refuge in the leaf-base and young shoots of palm trees, or in bushes and camel-scrub.

An interesting example of the periodic movement of animals is recorded from a small island off the coast of Ceylon. This is inhabited at night by crows which roost in the trees. In the morning they fly to the mainland to feed, and their place is taken by fruit-eating bats or 'flying foxes' which fly to the island at dawn.

The nocturnal habit confers several advantages: enemies are more easily avoided and food more easily obtained; losses of moisture are reduced because the air at night is more nearly saturated; and intercommunication between members of the same species is facilitated, because odours are more readily conveyed. The carnivorous habit is less hazardous, too, at night, and competition is reduced when nocturnal carnivores, such as carabid beetles, prey on sleeping diurnal animals such as butterflies.

Not only are many factors involved; these factors may mutually influence one another. For example, whilst some nocturnal animals are never active in the daytime, others can alter their habits in different circumstances. Thus the African buffalo was very abundant until, in 1890, a terrible epidemic of rinderpest almost exterminated them in many places. Whereas previously the animals used to feed in herds in the open by day, the survivors retired to forests and dense swamps, feeding only at night. After a number of years, however, buffaloes increased considerably, and returned to their former diurnal behaviour. The change produced in the habits of game animals by hunting has already been mentioned.

Another factor to be considered is that although many species are active during a certain period of the day or night, and are quiescent for the remainder of the 24 hours, others exhibit differ-

ent kinds of activity at different times. Many normally diurnal birds migrate at night. The water-skater *Gerris rufoscutellatus* spends the daytime on the surface of ponds and streams, but flies from one locality to another at night, and many other aquatic insects such as beetles fly mostly at night, but swim actively throughout the day.

Truly social animals are seldom nocturnal, although nocturnal species may form aggregations for feeding or hibernation.

Nocturnal species are frequently primitive. This is particularly apparent in the case of insects, of which cockroaches, silver-fish, bristle-tails, stick-insects, and crickets are typically nocturnal forms: and the primitive species of more advanced orders are frequently nocturnal also. Primitive insects tend to have a lower metabolic rate than more advanced forms; they are usually slow and retiring, and live in cool, shady, or even dark environments. More highly evolved insects are quicker-moving, diurnal, and live in hot, light environments. There are many exceptions to this generalization, however, and a number of difficulties arise from it. For example, were the tropical conditions of the Mesozoic period in which the insects evolved cooler and darker than the tropics of today?

Again, many nocturnal animals such as elephants, hippopotamuses, tapirs, sloths, lemurs, and millipedes have long fossil records. There may therefore be an evolutionary aspect to the problem of nocturnalism, but studies of the physiology of vertebrate eyes suggest that diurnality and nocturnality come and go as mutation and ecological expedience direct. The few species which normally live in constant environments such as underground caves and wells are sometimes completely non-rhythmic.

PHYSIOLOGICAL RHYTHMS

From a physiological point of view, diurnal rhythms are basically of two kinds: *exogenous*, a direct response to environmental changes, that is to conditions external to the animal, and *endogenous*, rhythms which persist under constant conditions, and are so named as if originating in the animal itself. Examples of exogenous rhythms are afforded by the Springhaas

(*Pedetes*), a strictly nocturnal South African rodent which may come out of its burrow on dark and gloomy days, and the nocturnal leech *Glossiphonia*, which is non-rhythmic both in darkness and under continuous lighting, when it stays hidden under stones for several days. Endogenous rhythms are frequently correlated with environmental changes, although they are not a direct response to them. Thus if the cockroaches *Blatta* and *Periplaneta* are subjected to alternating 12-hour periods of light and dark, their activity is largely confined to the latter, although they may begin to move about shortly *before* the light is extinguished. This rhythm of activity persists for some days in continuous light or darkness, but eventually becomes evenly spread over the whole of the 24 hours, leaving only a slight residual rhythm which is unrelated to the previous conspicuous one. If the periods of light and dark are now doubled, the 24-hour cycle of activity and rest is still maintained, but the insects show an outburst of movement-activity either when the light is switched on, or when it is extinguished. In such a case, the environmental change is referred to as a 'clue'.

One of the best examples of an endogenous rhythm is afforded by a giant West African millipede (*Ophistreptus*). In this animal a 24-hour rhythm of movement has been shown by the writer to persist in constant temperature and light for periods up to 19 days! This rhythm is affected by fluctuating temperatures, but not by alternating light and darkness. Perhaps in their natural gloomy habitat in tropical forest, light is an insignificant environmental factor. Whatever the explanation in this case, two British millipede species *Blaniulus guttulatus* and *Oxidus* (*Paradesmus*) *gracilis* have a simple exogenous rhythm which is primarily a response to the normal cycle of light and darkness, although affected also by temperature.

The acquisition of an endogenous diurnal rhythm is also well illustrated by Congo Floor Maggots, which are the larvae of an African blowfly, *Auchmeromyia luteola*. The adult is a shade-loving fly which frequents human habitations where it lives on fallen fruit, fermenting garbage, excrement, and so forth. The eggs are usually laid in the dust in cracks of floors

of huts, and under the sleeping-mats of the African natives where the emerging larvae conceal themselves during the day. At night, however, when the inhabitants lie asleep on their mats, these maggots crawl up and suck their blood. The bite is apparently not painful, the Africans do not seem to mind it, and there are no pathological complications. The periodicity of these insects is correlated with a response to temperature. At night they are attracted to the warmth of human bodies, but after feeding the response becomes negative. The rhythm, which can easily be reversed experimentally, is acquired in the course of larval existence, and is regulated by the habits of the Africans who return to their houses to sleep at sunset. Once it has been acquired, however, the rhythm will persist under constant conditions for several days.

There are few endogenous rhythms outside the two major groups of vertebrates and arthropods. An exception is afforded by the sea-cucumber *Thyrone briareus*, in which the number of waves of contraction passing along the body increases between 4 p.m. and 2 a.m. in constant darkness. Innumerable metabolic and physiological rhythms exist in vertebrates, however. For example, the goldfish is decidedly diurnal, and shows a rhythm of activity and rest which persists for several days in continuous light and running water. Measurements of oxygen consumption indicate that goldfish in small groups are less active in the daytime than are single fishes or fish in larger groups. Over the entire 24 hours, however, oxygen-consumption per fish is the same, regardless of the size of the group to which it belongs. This is because the rhythm is less marked in small groups than it is in larger groups or single individuals. Young rats and mice born and reared in darkness show a 24-hour rhythm corresponding with that of their parents. Oxygen consumption and carbon dioxide production is rhythmic and independent of the food eaten. A diurnal spermatogenic cycle has been established in sparrows, daily deposits of dentine and enamel are laid down in the incisor teeth of rats, there is a diurnal ovulation cycle in chickens, and cycles of glycogen and urine production occur in man.

A somewhat different rhythm is seen in the moulting of insects, the hatching of their eggs and emergence of pupae. These proceedings usually occur at a definite time of day and will continue to do so for several generations under constant conditions. No doubt these, like the endogenous rhythms described above, are closely connected with metabolic rates.

Most animals, if thoroughly enough investigated, are found to show a combination of both types of rhythm, and these are designated *composite*. An example, which will be discussed later, is afforded by the woodlouse, *Oniscus asellus*. This passes the day under leaves, logs, and stones, and shows a composite diurnal rhythm correlated with light and dark.

An alternative and perhaps more fundamental interpretation has been suggested in which exogenous and endogenous activity are separated on the basis of velocity. In exogenous rhythms the velocity of an innate physiological rhythm is affected by environmental factors, whilst endogenous rhythms are those which endure without such modifications.

Diurnal rhythms are of importance in the ecology of all except the least rhythmic species. They have been investigated in a large number of animals, and many references will be found in one or other of the reviews cited at the end of this article. No doubt in most instances the stimulus or clue to which the animal responds is an increase or decrease in daylight, but drops in temperature are usually the stimulus to emerge in the case of nocturnal species which pass the day in burrows underground. Examples are afforded by the nocturnal green tiger-beetle *Cicindela* and the Californian side-winding rattle-snake *Crotalus cerastes*, which inhabits the burrows of kangaroo-rats during the day.

Diurnal rhythms in animals can be studied in a variety of ways, such as by statistical analysis of the numbers abroad at different times of the day or night, by determination of the times of activity of individual animals in controlled environmental conditions, and by measurement of metabolic rate and other physiological changes. The writer is convinced that the most illuminating results are likely to be obtained by bringing as

many methods of investigation as possible to bear on the subject.

STATISTICAL INVESTIGATIONS

For practical reasons, much work on diurnal rhythms has been concerned with establishing the existence of 24-hour periodicity in animals of economic or medical importance. Three examples of this type of approach will therefore be considered before the results of laboratory experiments on the physiological nature of such rhythms are described.

Insect Flight Periodicity

The study of aphids has shown that there is a diurnal periodicity of flight and nocturnal quiescence which is reflected at heights of at least 2,000 feet above ground level. The greatest aerial densities, as shown by the catches of suction-traps at Rothamsted Experimental Station, occur between 9 a.m. and 9 p.m. G.M.T., and most species have two peaks of abundance each day.

Moonlight has a considerable influence on the activity of insects at night. For instance, the number of Noctuid moths attracted to light-traps is much reduced on nights of full moon, and there is a difference between the hours of flight of these Lepidoptera, according as the early or late portion of the night is moonlit.

From work of this kind we know that the majority of insects, as of other animals, possess a periodicity which responds to the alternation of day and night. Most Hymenoptera, including bees and wasps, are diurnal, as are predatory insects such as robber-flies (Asilidae), many tiger-beetles (Cicindelidae), scorpion-flies (Panorpidae) and dragon-flies (Odonata), while stag-beetles, chafers, termites, mosquitoes, bed-bugs, cockroaches, weevils, and many caterpillars are nocturnal. At first sight it may seem strange that insects which shun the daylight should be strongly attracted to artificial lights at night, but orientation responses may vary both with the strength of the stimulus and the physiological state of the recipient. But it has been found

that attack by Codling moth is reduced if apple-trees are illuminated at night, because oviposition is inhibited by light in this species.

Biting-cycles of Mosquitoes

The possibility that the noctural activity of mosquitoes may also be influenced by moonlight has received considerable attention in recent years, but the question seems to be a very complicated one. A knowledge of the biting-cycles of mosquitoes may have considerable economic and medical importance, particularly in respect to the epidemiology of malaria and yellow fever. For example, some species of *Anopheles* bite mostly at dusk, others in the middle of the night, others at dawn. Yet in many hospitals in Africa, mosquito nets are raised when the patients are awakened before sunrise. The inmates must therefore run an unnecessary risk of contracting infection. Troops are ordered to wear anti-mosquito clothing at dusk, yet are paraded unprotected in the morning, just when the biting-cycle of some species of mosquito reaches its peak.

Most experiments have been made by counting the number of mosquitoes attracted to human 'bait', and striking results have been obtained concerning biting activity in relation to vertical distribution. Although there is a sharp fall in biting activity with increase in height above the ground, the amount of feeding at ground level may vary considerably. It is evidently influenced by local conditions such as the presence of thick bush around camps and villages. In addition, low temperatures inhibit biting even at dusk, and the peak activity of crepuscular (twilight-active) species appears to be controlled by the light intensity at sunset.

Malarial parasites are probably particularly numerous in the blood-system at the peak times of the biting activity of mosquitoes.

Filarial Periodicity

Another periodicity which is important in medicine is that shown by the microfilaria larvae of *Wucheria bancrofti*, the

organism responsible for the terrible disease elephantiasis. The adults of this Nematode worm inhabit the lymphatic vessels of man, whereas the microfilaria remain in the blood-stream. They are rarely found in the blood-stream during the day, but towards evening appear in increasing numbers. By midnight there may be as many as 40 or 50 millions in the circulatory system of the patient. Their number decreases during the later part of the night, and by 9 o'clock in the morning they have again disappeared from the peripheral blood. This diurnal periodicity may be maintained with the utmost regularity for ten or more years. If the filarial patient is made to sleep during the day and remain awake at night, the rhythm of the microfilaria becomes reversed in 3 or 4 days. The presence of the parasites in the blood-stream is not a direct response to the sleepy state of the host, because they begin to appear several hours before the usual time of sleep, and do not disappear until after the usual time of waking.

Filarial periodicity has yet to be satisfactorily explained. It is obviously an adaptation to the biting-cycle of the mosquito *Culex fatigans*, the intermediate host of the parasite. It cannot be that the movement of microfilaria through the cutaneous blood-vessels is obstructed least when the body is fatigued – as some authors have suggested – because some varieties of *W. bancrofti* are non-periodic, and others show a reversed periodicity where the immediate host is a day-biting mosquito. Possibly the periodicity will eventually be satisfactorily explained by the cyclical parturition of the female filaria, with a daily destruction of all the embryos. Probably periodicity is a quality inherent in the microfilaria themselves.

LABORATORY EXPERIMENTS

The physiological necessity to animals of repose is familiar, if unexplained, and it is natural to expect that the alternation of periods of movement and rest should be correlated with the physical conditions of day and night. Diurnal rhythms of locomotion have been investigated in a number of species both visually, and by recording the movements of individual animals.

The apparatus used consists essentially of a box or cage pivoted along its median transverse line so that any movement of the animal within along the longitudinal axis causes the box to tip, and this is recorded by a needle writing on a smoked drum. Such an arrangement is known as an 'aktograph'.

Although in its normal habitat the grain beetle *Ptinus tectus* is non-rhythmic, in alternating light and darkness the greatest amount of activity occurs during the latter period. If the light is constant, however, and the temperature fluctuates, then the greatest activity occurs when the temperature is falling; and, after transfer to a constant temperature, this period still occurs at the same hour for several days. This investigation is important because it shows for the first time that an endogenous rhythm of movement-activity can be formed in a normally non-rhythmic species. Species shown to be non-rhythmic in nature, such as the cave-crayfish *Cambarus pellucidus*, the insect inhabitants of rotting logs, and so on, usually live in constant environments. As the environment of civilized man becomes more stable, he tends to be less clearly rhythmic in his activities, a point which will be discussed later.

Luminescence

Most of the animals that possess the power of producing light, other than those which inhabit the deep sea, do so only at night; and the diurnal rhythms of light production may persist for several days in constant darkness. An endogenous rhythm such as this is shown by the firefly *Photinus pyralis*, which begins flashing at the same time every evening if kept in constant darkness. It is possible that the rhythm of light-emission results from a rhythm of the central nervous system, since it is inhibited by removal of the brain or severing the ventral nerve cord. The fact that fireflies emit light continuously if placed in an atmosphere of pure oxygen suggests, however, that such rhythmical discharges from the central nervous system elicit the rhythmical physiological process of light-production. The fact that flashing can also be stimulated by light of a certain dim intensity shows that there may also be an exogenous component in the rhythm.

An endogenous rhythm of light production in marine Copepods, persisting for as long as twelve days in constant darkness, has been explained as a physiological cycle of secretion and discharge in certain cells, and a similar rhythm lasting for eight days in constant darkness has been demonstrated in Bermudan species of *Ptychodera* and *Glossobalanus*.

Many other observations on luminescence have been made, but not on its recurrence under constant experimental conditions. In every case it is very noticeable that light-production is inhibited by light, even at night.

Diurnal Migration of Eye Pigments

In many Noctuid and Geometrid moths, there is an endogenous diurnal rhythm of pigment migration in the compound eyes. This persists in complete darkness and does not seem to be correlated with diurnal fluctuations of temperature or humidity. Thus if the moth *Plusia gemma* is kept in darkness for several weeks, the eye appears to glow every night and appears black when briefly illuminated during the day.

Similarly there is a movement of the distal pigment cells in the eye of a Cuban shrimp *Macrobrachium*, but not of the proximal cells. Since the normal pigment movement at dawn or sunset is prevented by anaesthesia and by interrupting the blood supply to the eyes, it appears likely that the nervous system controls a metabolic periodicity by means of a hormone circulating in the blood-stream. The secretion of the 'sinus-gland' in the eye-stalk of the prawn *Palaemonetes* causes migration of pigment, and controls the diurnal rhythm of colour change (see below). Either there is a rhythmic cycle of secretion in this gland, or else a cycle in nervous and general metabolic activity which affects the gland and hence the eye-pigment. There is a detectable difference in the amount of hormone from the sinus-gland in the eye stalks of the light and dark adapted prawns.

When the lobster *Homarus* and the crayfish *Cambarus* were kept at constant temperature and in darkness (except for the brief periods when the observations were made) the diurnal rhythm of covering and uncovering the reflective layer of the

eye persisted for 5 months at 7° C. At 21° C. the rhythm disappeared after 4 months. Movements of the pigment in living Arthropods are very much easier to see than are the movements of the retinal pigment and rods and cones in vertebrates. Nevertheless a weak endogenous rhythm in darkness has been observed in the catfish *Ameiurus nebulosus*, lasting for two or three days.

Rhythmic Colour Change

Just as light-production and retinal changes are physiological processes easily observed, so also is colour change. In the lamprey, *Lampetra planeri*, there is a pronounced daily periodicity of colour-change – the animal is pale at night and dark during the day – but continuous light gives maximum darkening and halts the rhythm, whilst removal of the pituitary gland results in maximum paling. A similar exogenous rhythm occurs in many fishes, amphibia, and reptiles. In the common frog *Rana temporaria* it is correlated with complete dependence on illumination for activation of the pituitary. An endogenous rhythm of colour-change, without stimulus from light, is found in the South African clawed-toad, *Xenopus laevis*; in the Javanese flying frog *Rhacophorus reinwardti* the rhythm persists under continuous light stimulation.

Experiments on the larvae of *Salamandra maculosa* have established the existence of an endogenous periodicity persisting for up to 20 days in darkness and constant temperature: continuous light abolished the rhythm. Larvae raised in constant darkness and fluctuating temperature also showed a diurnal rhythm of colour change which was abolished when light and temperature acted in opposition to one another.

Many shrimps, prawns, and other Crustacea show exogenous, light-stimulated diurnal rhythms of colour-change because the expansion and contraction of their pigment-cells is influenced by hormones secreted in the eye-stalk under the stimulus of light. The marine woodlouse *Ligia baudiniana* is likewise pale during the night and dark in the daytime. Injection of aqueous extracts of the heads of light or of dark animals resulted in

paling, which suggests that diurnal pigmentary activity may not, as in the prawns, be due to a cycle of exhaustion and elaboration of endocrine material. Alternatively, in the secretory phase the hormone may be elaborated as fast as it is discharged.

In the Isopod crustacean *Idothea* an endogenous rhythm has been shown to last for 60 days, and in the 'stick insect' *Dixippus morosus,* which normally is dark at night and pale by day, the change will continue for several weeks in complete darkness. Here the effect of light is exerted solely through the eyes, and is determined by the contrast between the background and the surroundings.

DEGREE OF SPECIALIZATION

The question arises, how inherent are rhythms of 24 hours as opposed to, say, 15 or 18 hours? The buffaloes already cited were able to reverse their habits, and in man the degree of specialization to a nocturnal or diurnal existence is not particularly marked.

Like most young animals, the human baby is non-rhythmic during its first year of life, as parents are only too well aware, but gradually there is an increasing range and regularity of the 24-hour physiological rhythms of body-temperature, urine secretion and so on, as the child becomes progressively adapted to a normal rhythmic existence. Although as individuals we are markedly periodic, and those who have to work night shifts take a long time to become adjusted, nevertheless social communities as a whole are less periodic the larger they are. Thus the isolated family or hamlet may be completely rhythmic, but in a small town there is usually someone awake at night, if only the policeman. In larger cities there is a considerable degree of night life. An interesting parallel is afforded by ants and some other social animals.

Since in man it may take several days before the normal metabolic curve with its daytime peak is reversed, it has been suggested that greater individual and communal efficiency might result if night shifts were lengthened to several weeks or months, so that the rhythm might have time to become completely

reversed. Of course, means of recreation and other facilities would have to be available at night too, for the scheme to work.

On the other hand, albino rats kept in darkness show an endogenous rhythm up to 8 days. Young rats born and kept in darkness show 24-hour periodicity with a peak of activity which may be either in the day or night. If rats in which the peak is in the day are exposed to light for a single day, and then returned to continuous darkness, the peak of activity occurs thereafter during the night. A similar rhythm has been shown to persist for 24 days in a short-tailed vole, *Microtus,* and for over seven months in the American forest deermouse *Peromyscus,* although it got out of phase with day and night. Artificial illumination reversed the rhythm.

Attempts to establish artificial days of 16 hours in rats and mice have failed. Although the actual time of activity of a diurnal cycle may quite easily be shifted, the general plan of 24-hour periodicity seems to be innate. Although the writer has succeeded experimentally in establishing an 18-hour rhythm in the cockroach *Periplaneta americana,* this is somewhat feeble, and is less readily set up than a 24-hour rhythm.

The Time Sense of Social Insects

Some light is shed on this problem by investigations of the behaviour of social insects which have a time-sense, and can be trained to feed at regular intervals. In addition, there is an exogenous factor in the diurnal rhythm of bees, as was illustrated by a unique observation during a total eclipse of the sun in 1837. When the sun was darkened and the temperature dropped, the hive was quiet as at night. Full activity was resumed when the eclipse was over. Ants have a delicate temperature-sense by which their times of activity are regulated.

The time memory of bees and wasps is the expression of an endogenous rhythm based on innate metabolic periodicity. Insects were trained to visit their food places at certain times of the day, and then were given food containing certain drugs: the effect upon the time of subsequent visits was noted. It was found that euchinine delayed the maximal number of visits, and thyro-

iodine moved the visiting time forward. In the case of ants, 0.0008 per cent salicylic acid caused visits to food to take place at an earlier time of day than that to which the insects had been trained. A mixture of 0.001 per cent yellow phosphorus caused the average of visits to come as much as 6½ hours earlier than the training time, while arsenic trioxide in different dosage strengths altered the time of subsequent visits, either forward or backward, depending on its metabolic effect in different concentrations: probably both nervous conduction and temporal memory was affected.

The training of bees studied in relation to their normal feeding which is regulated by the availability of pollen and nectar, also suggests that their activity rhythms are correlated with deeper internal periodic patterns. Furthermore, ants, which may be non-rhythmic, have been trained to feed at several time intervals, whereas bees, which have a marked rhythm and are active only in the daytime, cannot be trained on other than a 24-hour basis.

Certain anaesthetics, and influences such as low temperature, carbon dioxide and quinine, have been shown to modify experimentally induced habits, whilst others, such as etherization, do not appreciably affect them. This again shows that time-memory is independent of external periodicities, and the 'clock' which determines the hour so accurately is again the rate of metabolism of the body. Presumably this time sense has been evolved in relation to the acquisition of food by social insects – bees, wasps, ants and termites – since it is not known in other groups of animals. In contrast, white rats have been shown to be able to estimate short time intervals only. Some humans are notoriously inefficient in this respect! *

In many species, however, there is a marked specialization for a nocturnal or diurnal habit. Diurnal animals have become adapted to meet, or to avoid, relatively high temperatures and

*It is not impossible that the ability of some people to judge time in their sleep and to awake with the utmost regularity may be correlated with the fact that the alpha rhythm of the brain is undisturbed whilst sleeping.

evaporation rates, bright light and decreased conductivity of the air for odours. Conversely, nocturnal species are adapted to decreased temperatures, high humidity, dim light and increased conductivity of the air for odours. It has been suggested that crepuscular species may be confined to the short periods of dusk or dawn, simply because they are relatively unspecialized in these respects; but this cannot apply to the Hawk-moths (*Sphingidae*) which are highly adapted to activity in twilight.

Scent and Hearing

In strictly nocturnal or diurnal species there is usually a refinement of one or more of the senses, and differences are of degree rather than of kind. Scent is perhaps most important at night for the congregation of individuals, sex attraction, following the tracks of prey, location of enemies, and so on; hearing and sound-production for communication, the detection of enemies or victims; light production for sex attraction and warning. Diurnal periodicity in sound production at dusk and dawn has been studied in the tropical rain-forest fauna of an island in the Panama Canal Zone. In this investigation there was found to be a correlation between sound-production by diurnal and nocturnal animals (cicadas, orthopterons, frogs, birds, and monkeys) and the physical influences, such as daylight, operating on them.

Sight

The sensitive elements in the retina of the vertebrate eye are of two kinds: rods and cones. Animals that are adapted for night vision have a particularly large number of rods. In man, the central part of the retina, the *fovea centralis* or 'yellow spot' on which objects are normally focused, consists almost entirely of cones. These are responsible for colour-vision and for detailed resolution. Towards the periphery of the retina the proportion of rods increases. That is why a very faint star can often be seen only when one looks slightly to one side of it, a principle frequently employed by astronomers. Colour-blindness in nocturnal animals such as mice is associated with lack of cones.

In the eyes of nocturnal animals a reflecting mechanism or

'tapetum' is usually present. Any light passing through the retina is reflected back so that there is a double chance for a given ray to stimulate a rod. The absence of a tapetum in owls illustrates the super-efficiency of the retinal rods through which little or no light passes. There are many kinds of tapetum, but reflexion is most frequently caused by fibrous tissue or guanin crystals. Associated with a very efficient retina as in the opossum, a tapetum results in such sensitivity that bright light is avoided at all times, even though a slit pupil very greatly reduces the amount of light penetrating to the retina, and ensures some protection by means of its fine control.

The larger terrestrial mammals – ungulates, elephants, lions, bears, and so on – have eyes which are not particularly specialized either for day or night vision. They have a tapetum and sufficient rods for reasonable visual sensitivity, and yet have many cones also. These, coupled with a large eyeball, ensure good resolution of detail. In consequence, vision is relatively good both at night and by day, even where senses of scent and hearing are acute, and the animals are not very markedly nocturnal or diurnal.

Nocturnal forms often have an iris with a vertical pupil aperture, and adaptations may also occur in the lens. For example, flying squirrels (*Glaucomys*) are nocturnal and have colourless lenses, whereas true squirrels (*Sciurus*) are diurnal and have yellow lenses. Both types inhabit the same forests in many localities, and their different periods of activity are related to the light-filtering properties of their eyes.

In diurnal eyes, intra-ocular oil droplets act in the same way as filters in photography. The red droplets in birds are effective at sunrise, but as the day wears on, the yellow and colourless droplets are most effective: at sunset the red oil droplets again come into operation. Thus late-rising birds such as hawks have fewer red droplets than song-birds. Red droplets enable turtles to see through the surface glare of a tropical sea.

Yellow filters cut out much of the violet light, and some blue. These are the colours responsible for most chromatic aberration. At the same time they enhance contrast, and do not impede

atural hues. A combination of a yellow diurnal lens or yellow roplets with a tapetum would result in poor vision at all times, nd does not occur in nature.

The majority of spiders have six or eight eyes, which appear o be of two kinds; some are black, others pearly-white. They ere at one time described as diurnal and nocturnal respectively, nd it was believed that one type was used in the daytime, the ther at night. The evidence for this is, however, slight. The chief isual organs of insects and crustaceans are the compound eyes, nade up of a number of transparent facets each with an elongated, light-sensitive structure beneath it. In diurnal insects images are formed by apposition – that is, the light passing through each facet falls only on the appropriate optic-rod or 'rhabdom'. In many nocturnal insects such as Noctuid moths, glow-worm beetles and the like, there is a special kind of compound eye that forms images by superposition. Each sensory rhabdom receives light rays not only through its own facet, but through neighbouring facets also. Between these two extremes there are many intermediate stages. Light and dark adaptation of compound eyes is achieved by migration of the pigment in each optical unit or 'ommatidium', so that the rhabdom is more or less shielded. Similarly migration of rods, cones and retinal pigment are of great importance in the lower vertebrates.

There is an Arabian fable concerning an argument between the horse and the lion as to which had the keener sight. The lion could distinguish a white pearl in milk on a dark night, but the horse could see a black pearl in a heap of coal in the day. The matter was submitted to arbitration, and the judges rightly pronounced in favour of the horse. No doubt they realized that the acuity of the cone vision of Ungulates is greater than that of the nocturnal rods and tapetum of Carnivores!

COMBINED METHODS

In this short account it has been possible to discuss only a few selected, but diverse examples of the kind of work that has been carried out on diurnal rhythms. It has already been suggested that a combination of the various methods of investigation is

likely to produce the most interesting and illuminating results, but in only a few cases has this been attempted. One of the best examples is afforded by planktonic animals which have been studied extensively on account of their economic importance as the staple food of fishes.

Diurnal Migration in Plankton

Many aquatic organisms show a periodic ascent and descent in the water in which they live. Several species of Crustacea, for example, migrate upwards at night, some at dusk or dawn, and a few during the day. Diurnal migration in plankton (the community of organisms inhabiting the surface waters of lakes and oceans) is a complex phenomenon. The behaviour of related species may vary considerably, and during its life-history the same animal even may show differing responses to identical stimuli.

Increase in temperature results in a tendency for the little 'water flea', *Daphnia pulex*, to swim downwards away from the light. The reverse occurs with decreasing temperature, and below 12° C *Daphnia* swims towards light of all intensities. Both increase in light and temperature result in the animals swimming downwards during the day. At night, however, decreased illumination and lower temperatures lead to the reverse response, and they move upwards. It is believed that reversal of the response to gravity with change in light-intensity is a predominant factor affecting diurnal migration in this species. In addition, however, *Daphnia* shows a diurnal periodicity in specific gravity. This is lowest in the early mornings, and in consequence *Daphnia* tends to float upwards from deeper water to the surface where it finds a plentiful supply of the algae on which it feeds. After a meal, during which its density increases it sinks slowly again. The diurnal rhythm therefore also depends in part upon variation in food supply, which itself is dependent upon light and temperature too. The periodic rise and fall towards and away from the surface waters is complicated in addition by the fact that after the brood pouch has emptied there is a temporary decrease in specific gravity.

Another fresh-water Crustacean, *Cyclops albidus*, normally swims downwards after exposure to light. The response becomes negative in darkness, however, so that the animals swim upwards at night, but after a time there is a spontaneous reversal again even if there is no change in external conditions. Similar changes occur in the aquatic larvae of the gnat, *Corethra*. *Arcella*, a little freshwater Rhizopod, shows a rhythm of sinking and floating, depending on its oxygen requirements; and innumerable other examples could be quoted.

Plankton migration is infinitely more important in the case of marine animals. Specimens of the Copepod *Acartia* taken from surface water initially exhibit different responses to light, gravity, and temperature than specimens from deeper water, although later all react in the same way. Here there is a physiological rhythm of negative and positive responses to gravity which are not directly related to changes in the environment.

If diurnal migration is correlated with light intensity as has been suggested, it is reasonable to expect that there should be a relationship between the transparency of the water and the depth to which the plankton migrates, and it has indeed been shown that planktonic animals are to be found deeper in the sea on sunny days, or when the water is particularly clear. Again some species congregate on the surface, especially on moonless nights, and fishermen shoot their drift-nets at night because this is when the herrings come to the surface for their food.

In conclusion it must be emphasized that each species has its own way of performing vertical movements, and no general explanation is possible. Changes in response to gravity with light and temperature are probably not sufficiently widespread to be considered of universal significance.

Tidal Periodicity

Most sea-shore animals have a marked rhythm correlated with the state of the tide. At low tide they are inactive, and they are activated by the force of the waves. In the laboratory a rhythm of sensitivity to mechanical shaking persists in periwinkles (*Littorina*) for some days. The little green flatworms

Convoluta roscoffensis, which owe their colour to the presence of symbiotic algae in their tissues, come to the surface of the sand in great numbers during high tide. At low tide, they descend into the sand as it dries. If placed in an aquarium, this rhythm will persist for as long as 8 days in close co-ordination with the changes of tide in the ocean. Tidal rhythms are related to the lunar day of 24 hours, 50 minutes, and so, strictly speaking, do not come within the scope of this essay. But obviously the time of day of exposure at low tide will have a considerable effect upon the littoral fauna.

Diurnal Rhythms in Woodlice

Students of the behaviour and orientation of animals cannot really afford to neglect diurnal rhythms, since the responses of an animal to its environment may vary at different times. For example, during the active phase of its diurnal periodicity the well-known responses of *Blatta orientalis* to light and temperature seem to disappear to some extent, and the insect becomes indifferent to its surroundings. The pill-woodlouse *Armadillidium vulgare* is more resistant to desiccation than other species of terrestrial Isopoda, and is chiefly active in the morning. This is probably correlated with the fact that in this species some individuals walk towards the light when the temperature rises.

The writer has recently investigated diurnal changes in the physiological responses of the common woodlouse *Oniscus asellus*. This nocturnal species has a composite diurnal rhythm of movement correlated primarily with alternating light and darkness, and not with fluctuating temperature and humidity.

The intensity of the humidity responses of the woodlice is less in darkness than in light, and is less still in the nocturnal phase; it increases with desiccation. Movement away from light is more marked in animals which have been in darkness for some time, and is still shown in dry air, whereas control animals move towards the light when humidity is sufficiently reduced. Finally the active phase is correlated with increased sensitivity to external conditions such as the carbon dioxide content of the air.

These experimental results are readily correlated with the nocturnal ecology of the species. Woodlice are often to be seen wandering in dry places at night, but they spend the daytime under stones, logs, and in other damp, dark situations where environmental factors are relatively constant. If nightfall only involved a reduction of the conditions which normally operate to restrict them to their daytime habitats, it seems unlikely that they would have any occasion to leave them. The endogenous component of the diurnal rhythm, however, will engender locomotory activity in some individuals at least, at nightfall when they are especially sensitive and easily disturbed. Even if the majority are exposed to daylight only occasionally, this may be sufficient to maintain their periodicity and to keep it in phase with the 24-hour cycle. The decrease in the intensity of the humidity response enables them to walk in dry places where they are never to be found during the day, and the increased photo-negative response in darkness ensures that they get under cover promptly at daybreak, and thus avoid many potential predators. On the other hand, if their daytime habitat should dry up, the woodlice are not restrained there until they die of desiccation, since they become photo-positive in dry air and thus are able to wander in the light until they find some other damp hiding place, and again become photo-negative.

Innumerable other questions remain to be investigated too. No doubt different mechanisms are responsible for the rhythm of colour change in a prawn and the nocturnal wakening of a bat, but are there several timing mechanisms in the same animal? No evidence has yet been obtained of cyclical secretory activity of the endocrine glands or of a nervous-endocrine-metabolic sequence controlling diurnal rhythms, but the effect of anaesthesia suggests that the nervous system plays an important rôle in the maintenance of endogenous periodicity. Nor is it known whether a single process is involved or a whole cycle of processes as in the case of the vertebrate sexual cycle. In conclusion it is emphasized that this article has been concerned only with a few aspects of one of the many rhythms to which nature is heir.

FURTHER READING

The following are recent review articles; the first, particularly, gives extensive further references:

Calhoun, J. B. 'Twenty-four hour periodicities in the animal kingdom'. *Journal of the Tennessee Academy of Science*, **19**, 179 (1944); **20**, 228 (1945); **21**, 208 (1946).

Cloudsley-Thompson, J. L. 'Diurnal Rhythms'. *Transactions of the IXth International Congress of Entomology*, Vol. 1, 305-10 (1952).

Kikuchi, K. 'Diurnal migrations in plankton crustacea'. *Quarterly Review of Biology*, **5**, 189 (1930).

Park, O. 'Nocturnalism—the development of a problem'. *Ecological Monographs*, **10**, 485 (1940). Chapter 20 in Allee, W. C., *et al. Principles of Animal Ecology*. Philadelphia, U.S.A., 1949.

Uvarov, B. P. 'Insects and Climate'. *Transactions of the Royal Entomological Society*, **79**, 1 (1931).

Walls, G. L. *The Vertebrate Eye and its Adaptative Radiation*. Cranbrook Press, Bloomfield Hills, Michigan, U.S.A., 1942.

Welsh, J. H. 'Diurnal Rhythms'. *Quarterly Review of Biology*, **13**, 123 (1938).

WHAT IS A METAL?

J. CROWTHER

MOST people have a fairly clear idea of what they expect a metal to be like. It has a characteristic 'metallic lustre' when polished or freshly cut; it is a good conductor of heat and tends to feel cold to the touch; it conducts electricity well; if sufficiently massive it rings when struck; and it can often be bent or hammered into shape. Metal articles are generally weighty, and it is well known that the common metals, at least, melt when sufficiently heated. Such properties are fairly general amongst the metals made familiar by long usage, but it is not difficult to show that they are not all exclusive to metals or that many metals do not possess one or other of them. Glass rings when struck, whilst lead and tin do not; cast iron is notoriously brittle; aluminium and magnesium are of low density; graphite is a good conductor and has a metallic sheen when polished. If more uncommon materials are considered, it becomes more and more difficult to decide from such simple tests whether they are metallic or not. Further, ordinary experience does not allow us to distinguish between a 'metal', in its strictest sense as a metallic chemical element, and an 'alloy', an intimate mixture of two or more metallic elements which possesses obvious metallic properties.

Of the 92 elements, from hydrogen to uranium, 68 are generally acknowledged as metals, 19 as definitely non-metals, and the remaining 5 are somewhat difficult to classify. Some of these 68 are of very common occurrence in the earth's crust, some are extremely rare, and a glance at Table I will show that the familiar ones are not necessarily the most common. Thus titanium,* which is only just becoming commercially available, is far more abundant than copper, which has been in use for

*See, for example, Fox, F. A. *Science News 26*, 98 (1952).

more than 5,000 years. Less than half of them have any significant industrial use. The quantity of iron produced far exceeds that of all other metals put together, being followed, at a considerable distance, by manganese, copper, aluminium, and zinc.

TABLE I: *Percentages of metallic and non-metallic elements in the earth's crust, oceans and atmosphere (Smithells)*

Metallic elements		Non-metallic elements	
Aluminium	7.45	Oxygen	49.13
Iron	4.20	Silicon	26.00
Calcium	3.25	Hydrogen	1.00
Sodium	2.40	Carbon	0.35
Potassium	2.35	Chlorine	0.20
Magnesium	2.35	Phosphorus	0.12
Titanium	0.61	Sulphur	0.10
Manganese	0.10	Fluorine	0.08
Barium	0.05	Nitrogen	0.04
Strontium	0.04		
Chromium	0.03		
Zirconium	0.025		
Vanadium	0.02	The remaining 67 elements make up only 0.15 per cent between them	
Nickel	0.02		
Zinc	0.02		
Copper	0.01		

Very few metals – the noble metals, gold, platinum and its allies, together with some silver and copper – exist in the earth's crust in metallic form. All the rest are present as chemical compounds ('ores'), oxides, sulphides, sulphates, carbonates, etc., which usually bear no resemblance to metals and from which the metal is obtained by a 'reduction' process of some kind, whose difficulty increases with the chemical activity of the metal. Once won from its ore, the metal continuously tries to return to the combined state, as the common experience of 'rusting' demonstrates.

Stone Age man probably knew metals only as occasional nug-

gets of the precious metals and prized them, if at all, for their abnormal weight or for their malleability. Nevertheless, he used many natural metal compounds, malachite (green copper carbonate) and stibnite (black antimony sulphide) being used as cosmetic pigments in predynastic Egypt (about 3500 B.C.). Some of these compounds are readily reduced to metal in a charcoal fire and metals were probably thus produced by accident: copper first became important, then bronze (a copper-tin alloy), and finally iron. In Roman times, about 200 B.C., seven metals were fairly well known and used in quantity – iron, copper, lead, tin, mercury, silver, and gold – and these seven metals have ever since largely determined our ideas about what a metal should be like. They are all of high density (Table II), lustrous and cold to the touch. Apart from mercury, which is liquid at ordinary temperatures, they are malleable (silver and gold extremely so), they harden when hammered and soften again when heated. They are fusible (although at that time and for a long time afterwards iron was not normally melted), and most of them alloy readily with each other when melted.

TABLE II: *Specific gravities and melting points of seven metals in use about 200 B.C.*

Element	Specific gravity	Melting point (° C.)
Iron	7.86	1528
Copper	8.93	1083
Lead	11.34	327
Tin	7.28	232
Mercury	13.55	−39
Silver	10.50	960
Gold	19.3	1063

Alloys were made by melting together two or more of these metals (e.g. copper and tin to make bronze, tin and lead to make low-melting solders), or by smelting mixed ores. The ancients

had no means of distinguishing an alloy from a pure metal and probably often smelted, by accident, surprising and unrepeatable alloys or even new metals not recognized as such until many centuries later. Metals were probably only reproducibly obtained from particular sources of ore and hence often carried their place of origin as a kind of hall-mark. The Romans had at least three kinds of 'lead' according to Pliny – *plumbum nigrum* (ordinary lead), *plumbum candidum* (tin, from Britain), and *plumbum cinereum* (bismuth), whilst copper was *aes cyprium* or bronze from Cyprus. Fifteen hundred years later nickel and cobalt ores were still confusing the copper smelters of Germany, and got their names as creatures of the devil (Nickel and Kobold).

By the eighteenth century, many minerals were recognized as being compounds of metals hitherto unknown and one after the other these metals were obtained in recognizable metallic form, although often with great difficulty. The ancient criteria were, however, still applied and sometimes led to doubts about the new material being truly a metal or not. For example, to Henckel (1743) zinc was 'a metal with regard to its consistency, lustre, specific gravity, tenacity, and mercurial fluidity in the fire, but also not a metal with respect to its flammability and complete combustibility wherein it is entirely different from all other metals'; and a crude cobalt also was 'a half metal – resembling metal in colour, weight, and form, but not malleable'.

All smelting hitherto had been based on reduction of ores by heating with some form of carbon, but in the early nineteenth century Davy and Faraday first used the new 'voltaic' electricity to isolate the alkali metals, sodium and potassium, from their fused compounds; these two metals differed in many ways from the traditional ones, being very soft, lighter than water, and highly reactive. They were used in turn to reduce compounds of other unknown metals, calcium, magnesium, aluminium, etc., which were too powerfully combined to be reduced by carbon, but all these remained chemical curiosities until cheap electric power became available. New methods of attaining high temperatures allowed the isolation in impure form of a further type

of metal – vanadium, niobium, tantalum, etc. – with still more unusual properties, including great chemical stability at ordinary temperatures but such affinity at high temperatures for oxygen, nitrogen, carbon, and even the best refractories that it is only in the last decade or so that these metals have been obtained in a form sufficiently pure for us to be really sure of their properties.

The nineteenth-century concept of an element was essentially a chemical one, and so it is not surprising that in classifying the elements into 'metallic' and 'non-metallic' the chemist used as criteria certain differences in chemical properties, while accepting such of the older, physical, criteria as he found convenient. The chemist's ideal metal had the following properties:

(1) It combined with oxygen to give a basic oxide which was soluble in acids to form salts. (The oxides of most non-metals *are* acids.)

(2) Its salts when ionized in solution gave an electro-positive metal ion, which in many cases could be plated out on to the negative pole of an electrolytic cell. (Non-metals go into the electronegative ion.)

(3) It gave a chloride which was stable in water. (Most non-metal chlorides are decomposed by water – some violently so.)

The chemist's ideal metal is obviously one of the alkali metals, like sodium or potassium, but to the ancient metallurgist, or for that matter the modern one, such a substance would seem a mere caricature of a metal. Fusible, certainly (melting points: Na 97° C., K 63° C.), but not miscible with most other metals; malleable, but soft enough to be easily cut with a knife. The alkali metals are of very low density; lithium, sodium, and potassium float on water, and the lightest of all, lithium (specific gravity 0.57) is far less dense than most woods. They are much too soft to feel metallic and tarnish so rapidly in air that the metallic lustre of a freshly cut surface is quite evanescent; potassium catches fire in contact with water.

H. R. Paneth has discussed (in *Science News 24*) how chemical knowledge was rationalized during the last eighty years in the Periodic Classification of the Elements, and a version of the

Periodic Table, somewhat similar to that accompanying his article, is given in Table III. The heavier line indicates the boundary between metallic and non-metallic elements. The inert gases, beginning with helium (He), are shown on both the right and left sides of the table in such a way as to suggest that its real shape should be a continuous spiral. The apparent gaps between H and He, Be and B, Mg and Al, are inherent in the method of presentation and should be ignored.

The alkali metals appear as group I in the table, one column to the right of the inert gases. On the opposite side of the table to the alkali metals (or just beyond the inert gases is another way of looking at it) are the ideal non-metals – the halogens; fluorine, chlorine, bromine, and iodine. Throughout the table there is a gradual transition between these two extremes. In the first two horizontal periods named above, there is no doubt as to where the change from metal to non-metal takes place; it is between the second and third elements, beryllium and boron, in the first row, and between the third and fourth elements, aluminium and silicon, in the second row. A glance at Table III will show that metallic elements comprise an increasing proportion of each succeeding period, and in general metallic properties become more pronounced as we move to the lower atomic weight in each horizontal period and to higher atomic weight in each vertical group.* The reason for this will be discussed later.

The border line between metal and non-metal is never quite sharp. Beryllium and aluminium, although undoubted metals physically, give oxides which can show acidic as well as basic properties and their chlorides are not quite happy in aqueous solution. Taking a vertical group as an example, nitrogen and phosphorus are unquestionably non-metallic in appearance and properties; arsenic looks like a metal but behaves more like a non-metal; bismuth is predominantly metallic but is still very brittle and its chloride is not really stable in water. There is a distinct metallic appearance about all the border line non-metals, boron, silicon, germanium, arsenic, and tellurium, and even in one form of carbon, graphite. The chemical behaviour

*This is the diagonal rule, mentioned in Dr Paneth's article.

of the elements in the vertical groups from titanium to manganese is very complex and so unlike that of the typical alkali metal that one feels that the scrupulous chemist would only call them metals on physical grounds! The position of hydrogen was for a long time a subject of acute controversy. Its oxide (water) is completely neutral, but it has a stable chloride, nitrate, etc., like many undoubted metals; and, when these are dissolved in water, hydrogen forms the positive ion. Many chemists claimed that when ultimately it was obtained in the solid form it would prove to be physically a metal, but that was not so; there is, however, no real anomaly in putting it at the head of the alkali metal group in Table III.

The next development in ideas about metals came from the physicists. To the nineteenth-century chemist, an atom was the ultimate particle of matter into which an element could be divided and still retain its identity, and was, in fact, little more than an abstraction permitting a relatively simple explanation of observed chemical behaviour. The atoms of any one chemical element were characterized by a fixed weight relative to a standard (an atom of oxygen) and a certain 'valency' (the power of combining with atoms of other elements in a way that could be visualized as being by a fixed number of links).

In the last fifty years the physicist has established that an atom has a nucleus and a number of orbital electrons. In the inert gases, helium, neon, etc., the numbers of electrons are such as to form stable shells (see Paneth's article for details) and, to a first approximation, the chemical properties of an element are largely determined by the extent to which its array of electrons departs from an ideal stable shell. Thus, the alkali metals, one group to the right of the inert gases, have one extra electron, which they readily lose, whilst the halogens on the left are one electron short, a deficiency which they would like to make good. An atom of chlorine, Cl, readily accepts an electron from an atom of sodium, Na, to give a stable compound, NaCl. The movement of the electron with its negative charge converts the originally neutral atoms into electrically charged ions, Cl^- and Na^+, the existence of which can be demonstrated in solution.

TABLE III: *The Periodic Table of the Chemical Elements*

	Group																	
0	1	2								3	4	5	6	7	0
	1 H 1·0080														2 He 4·003
He	3 Li 6·940	4 Be 9·02								5 B 10·82	6 C 12·010	7 N 14·008	8 O 16·0000	9 F 19·00	10 Ne 20·183
Ne	11 Na 22·997	12 Mg 24·32	(Al)							13 Al 26·97	14 Si 28·06	15 P 30·98	16 S 32·066	17 Cl 35·457	18 A 39·944
A	19 K 39·096	20 Ca 40·08	21 Sc 45·10	22 Ti 47·90	23 V 50·95	24 Cr 52·01	25 Mn 54·93	26 Fe 55·85	27 Co 58·94	28 Ni 58·69	29 Cu 63·54	30 Zn 65·38	31 Ga 69·72	32 Ge 72·60	33 As 74·91	34 Se 78·96	35 Br 79·916	36 Kr 83·7
Kr	37 Rb 85·48	38 Sr 87·63	39 Y 88·92	40 Zr 91·22	41 Nb 92·91	42 Mo 95·95	43 Tc 99	44 Ru 101·7	45 Rh 102·91	46 Pd 106·7	47 Ag 107·880	48 Cd 112·41	49 In 114·76	50 Sn 118·70	51 Sb 121·76	52 Te 127·61	53 I 126·92	54 Xe 131·3
Xe	55 Cs 132·91	56 Ba 137·36	RE	72 Hf 178·6	73 Ta 180·88	74 W 183·92	75 Re 186·31	76 Os 190·2	77 Ir 193·1	78 Pt 195·23	79 Au 197·2	80 Hg 200·61	81 Tl 204·39	82 Pb 207·21	83 Bi 209·00	84 Po 210	85 At 211	86 Rn 222
Rn	87 Fr 223	88 Ra 226·05	89 Ac 227	90 Th 232·12	91 Pa 231	92 U 238·07	93 Np 237	94 Pu 239	95 Am 241	96 Cm 242	97 Bk 243	98 Cf 244						

Elements below and to the left of the heavy line are classed as metals; those above it and to the right, as non-metals. The inert-gas elements, shown at the right of the table, are repeated one period lower at the left to suggest that the real shape of the table should be a continuous spiral. RE = lanthanum and 14 rare-earth metals. The numbers to the left of the symbols of the chemical elements are atomic numbers. The numbers below the symbols are atomic weights relative to oxygen = 16.

The characteristic chemical properties of the chemist's ideal metal can be regarded as the result of its having a single valency electron which it readily loses to form a monovalent ion. Similarly the characteristic physical properties of a metal can be largely explained as a result of its 'metallic' crystal structure, which again is a consequence of the ready formation of an ion. A metal, therefore, may be looked upon as an element whose atoms readily form electropositive ions, and when this is conceded it is obvious that metallic elements will be found most readily at the beginning of any period in the Periodic Classification; that is towards the left of the table as shown. In addition, the heavier the atoms of an element, the more numerous are the orbital electrons which balance the increased nuclear change; thus the farther away from the nucleus will be the valency electrons, which are the outermost. Hence in any vertical group the heavier elements will be more easily ionized than the lighter, thus explaining the steadily increasing preponderance of metals amongst the heavier elements.

In the table of the natural known elements, a definite metallic tendency has encroached as far as group 6 in some of the properties of tellurium, and there is even a suspicion in the metallic lustre of iodine, the heaviest known member of the halogen family – the family that begins with fluorine, the supreme example of a non-metal. A by-product of recent 'atomic energy' investigations is astatine, a previously unknown halogen heavier than iodine; this is sufficiently 'metallic' to allow deposition in certain circumstances on the negative pole of an electrolytic cell, and appears to complete the process of confining the real non-metallic elements to the top right-hand corner of the Periodic Table.

A more recent extension of these ideas deals with the structure of solid materials. Solid sodium chloride is a regular pattern of electrically charged sodium and chlorine ions arranged alternately in three dimensions. Such an arrangement is stable and rigid, since deformation or fracture involves the separation of electrical charges; the bonding is known as 'ionic'. Carbon atoms which can give or accept four electrons can form a solid by each

atom sharing its electrons with four others, thus simulating eight electrons per atom; the four other carbon atoms are arranged symmetrically in space as if at the four corners of a tetrahedron and the whole structure is that of the hard, infusible but brittle diamond with 'co-ordinate' bonds.

Metallic sodium, however, makes a solid aggregate in a very different way. The sodium atoms shed their odd electrons and settle themselves into a compact arrangement of roughly equidistant positive ions with a kind of mobile cement of valency electrons keeping them together and maintaining overall electrical neutrality. The actual pattern is shown in Figure 1 and is known as 'body-centred cubic' since its smallest unit can be regarded as a cube with ions at each corner and one in the middle. An even more compact type is shown by copper; it is known as 'face-centred cubic' (Figure 2), and its basic unit is a

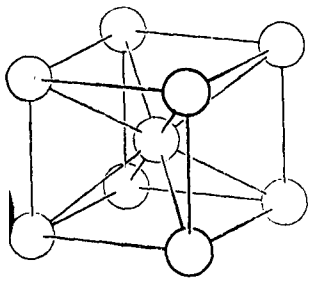


Fig. 1. Body-centred cubic.

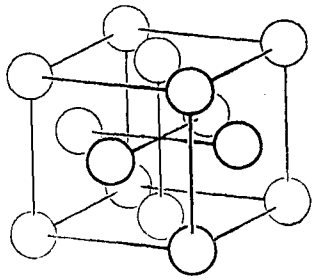


Fig. 2. Face-centred cubic.

These are the smallest elements of pattern in the lattice; they are repeated many thousands of times in all directions in any actual metal crystal. The use of heavier lines for some of the ions is to help in obtaining a three-dimensional effect.)

cube with ions at each corner and at the centre of each face. This is one of the ways in which uniform spheres can be most closely packed and is readily demonstrated with actual spheres such as table-tennis balls. A third type, shown by zinc, is called 'hexagonal close packed' (Figure 3) and has the same density of

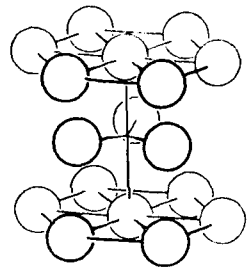


Fig. 3. Hexagonal close packed.

These three 'metallic' crystal structures have properties ver different from those of sodium chloride or of carbon. First o all, the valency electrons are not bound to any one atom but ar free to drift throughout the crystal. In an isolated crystal thi drift is random, but if an electric field is applied the drift be comes unidirectional and constitutes an ordered flow of negative electricity – that is, an electric current. Hence the typical meta is an electrical conductor. Good electrical conductivity brings in its train good thermal conductivity, since the electrons can also transmit thermal vibrations efficiently; and in addition high reflectivity, since a conducting sheet will reflect the short-wavelength electromagnetic radiations which constitute light in much the same way that a very much larger sheet of metal reflects the long wavelength electromagnetic radiations which are radio waves. It is no coincidence that polished silver is the best light reflector known as well as being the best electrical conductor.

Secondly, since a crystal lattice composed of close-packed atoms gives the greatest number of atoms possible in a given volume, metals tend to have a higher density than non-metals of similar atomic weight. The real reason, however, for the idea of high density as an attribute of metals is that the typical and best-known solid non-metals (e.g. carbon and sulphur) are of low atomic weight while the older metals (e.g. copper and lead) are of much higher atomic weight.

A third feature is perhaps the most important of all, the fact

that a metal crystal can be plastically deformed without fracture. One layer of atoms can be slid over the next and, in a perfect crystal, movement through one atomic diameter, or two, or any whole multiple of a diameter would leave the structure indistinguishable from the original except for overlapping edges. The way in which deformation occurs cannot be discussed here, but it can be said that the face-centred cubic structure has many possible directions of atom movement and characterizes the most ductile metals like silver, copper, and gold, known for thousands of years for their malleability. The body-centred lattice typical of iron has somewhat more restricted flow, and the hexagonal lattice of zinc and magnesium flows easily only in one plane of the crystal. These three kinds of structure cover most of the metals (Table IV) and there is a surprisingly even distribution between the three. It will be noted that antimony

TABLE IV: *Distribution of the three main types of metal structures*

Face-centred cubic	Body-centred cubic	Hexagonal close-packed
Aluminium	Barium	Beryllium
Calcium	Columbium	Cadmium
Copper	Chromium	Cobalt
Gold	Iron	Hafnium
Iridium	Lithium	Lanthanum
Lead	Molybdenum	Magnesium
Nickel	Potassium	Osmium
Palladium	Rubidium	Rhenium
Platinum	Sodium	Ruthenium
Rhodium	Tantalum	Thallium
Scandium	Tungsten	Titanium
Silver	Vanadium	Yttrium
Strontium		Zirconium
Thorium		Zinc

*Excluding rare earth metals, which are mostly hexagonal, like lanthanum.

and bismuth do not appear in the table. They do not show a typical metal lattice, but one similar to that of the non-metal arsenic, which explains their brittleness and relatively low electrical conductivity. Another omission is of the well-known metal, tin. Its normal structure can be regarded as a somewhat distorted cubic one, but at low temperatures it can change into a diamond-like structure (grey tin) which is brittle and of low conductivity, reflecting the somewhat ambiguous chemical behaviour of the metal.

So far we have considered the properties of the metallic elements, but so many metallic materials are alloys and not pure metals that it would be unfair to leave out their consideration entirely. Any real theoretical comparison of alloys and their structures is very modern and in an early stage of development, but useful ideas have already appeared.

As the ancients knew, the melting together of two metals usually produces a homogeneous liquid and, on freezing the liquid, an apparently homogeneous solid. On a microscopic scale, however, we now know that the solids are generally not homogeneous and that several types are possible. In a simple 'eutectiferous' system the liquid breaks down into the elementary metals, and the resulting solid is a very fine pattern of intimately mixed particles of the individual metals. Such an alloy, as might be expected, tends to have properties rather like those of the constituent metals – a lead-tin alloy (solder) is basically very like the two pure metals, although somewhat harder and of lower melting point. A second type occurs when atoms of one may be freely replaced in its crystal lattice by atoms of the other metal – e.g. copper and nickel, silver and gold. Here the alloy may be considerably stronger and of lower electrical conductivity than the parent metals, but as in the first case the differences are in degree rather than in kind.

In a third type, however, the metal atoms associate to form 'compounds'. When these were first recognized they were regarded as being similar to chemical compounds, obeying the ordinary laws of chemical valency, but the past 20 years has shown that new concepts of structure are required to explain the

existence of most of them. Often chemical valencies are obeyed, as, for example, in Mg_2Si, Mg_2Ge, Mg_2Sn, Mg_2Pb, the series of compounds formed by magnesium with elements of group 4, but normal valencies cannot be invoked to explain $Cu_{31}Sn_8$ or $Al_4CuMg_5Si_4$.

Hume-Rothery showed that in the analogous alloy systems copper-zinc, copper-tin, silver-zinc, silver-tin, the addition of increasing amounts of the second element (zinc or tin) to the first (copper or silver) produced first of all a solid solution in the parent face-centred cubic lattice with slight change in properties (an 'alpha' solid solution) but then a second, 'beta', solid solution based on a different kind of lattice which was body-centred cubic.

The fundamental constitution of these beta solid solutions is determined by chemical valency, but in a peculiar way – that the ratio of the total number of valency electrons to the total number of atoms is 3 : 2 – e.g. in the copper-zinc system the formula is CuZn, giving two atoms and three electrons (one from the copper and two from the zinc), whereas in the copper-tin system the corresponding compound is Cu_5Sn (six atoms and nine electrons since the valency of tin is 4). Similar compounds in other systems are AgZn, AgCd, AuZn, AuCd and Cu_3Al. All these, although compounds, have obvious metallic characteristics and are malleable, at least at high temperature; in fact many ordinary kinds of brass are of the beta type when heated and are made especially for the ease with which they may be fashioned by hot stamping into everyday articles like water taps and door knobs.

With further addition of the second element in these systems a 'gamma' structure results, which is complex, hard and brittle The electron/atom ratio here is 21 : 13, which gives some very peculiar formulae, Cu_5Zn_8, Cu_5Cd_8, Cu_9Al_4, $Cu_{31}Sn_8$. These alloys are still lustrous and reasonable conductors of electricity but have lost practically all metallic ductility.

Many metals are prolific compound formers, and since most of these compounds are completely lacking in ductility, although not always hard, very wide ranges of alloy compositions are

completely unusable because they contain a predominance of such compounds, which can often be readily powdered with a hammer. For example, copper and aluminium are both highly ductile metals and alloy in all proportions, but alloys of copper with from about 15 per cent to 75 per cent of aluminium are without any ductility whatsoever.

A small admixture of such compounds is often highly beneficial in increasing strength without sacrificing too much ductility; particularly in cases where the compound can be made to dissolve on heating to form a ductile solid solution and then separate in submicroscopic form on cooling to impart great strength to the alloy. The well-known aircraft alloy, duralumin, was the first industrial alloy to make use of this phenomenon, and to-day this method of hardening is used in hundreds of ferrous and non-ferrous alloys from lead cable sheathing to gas-turbine blading. In all these, however, the compound is but a small proportion of the alloy and the major part is of a metal or solid solution with enough ductility to permit its fabrication into a useful shape.

Brittle alloys were not unknown to the ancients. If they added too much tin to copper in making bronze, or too much carbon to iron in making steel, they obtained a quite unworkable material which often no doubt caused them bewilderment as well as discomfiture. The search for stronger alloys is always leading the metallurgist to the borderline where decreased ductility sets a limit to the strength attainable, although nowadays the reasons are somewhat better understood. So far, his engineering materials remain definitely metallic (except when he brings in carbon or nitrogen as an alloying addition), but there are occasions when inadequate control can almost cause the frontier to be crossed. For example, in protecting iron with a layer of aluminium, or tin, or zinc, too much diffusion of one metal into the other produces an intermediate layer of metallic compounds, which is weak by reason of its intense brittleness.

CONCLUSION

The ancients discovered and learnt to use seven metallic elements. To-day science knows ten times as many elements and thousands of metallic alloys, some of which the ancients would think very queer metals indeed. The chemists and the physicists have their own special conceptions, but the common man of our day still thinks of a metal in the same general terms as did the Romans, as something that can be melted and cast, hammered hot or cold, and that looks and feels like a metal.

FURTHER READING

A simply written introduction is:

Alexander, W., and Street, A. *Metals in the Service of Man* (Pelican A125). Harmondsworth, Penguin Books, 1944.

For an introduction to physical structure and properties at a relatively advanced level, *see*:

Hume-Rothery, W. *The Structure of Metals and Alloys*. London Institute of Metals Monograph No. 1, 2nd ed., 1944.

On the discovery and early uses of metals, the inevitable starting-point is still:

Partington, J. R. *The Origins and Development of Applied Chemistry*. London, Longmans, Green and Co., 1935.

RESEARCH REPORT

A. W. HASLETT

PROTEIN, NUCLEIC ACID, AND ENZYMES

In a discussion held lately by the Royal Society on the functi and location of enzymes in cells, Professor F. Binkley, of Emo University, Georgia, described experiments with potentia wide implications in relation both to enzyme activity and t building up of proteins. Although his conclusions under the two headings are of a different order of certainty, it is convenie to state them together in the first place. These conclusions a that the dipeptidases, enzymes plausibly considered to cont bute to protein synthesis, do not themselves depend on prote for their activity, but are essentially pentose nucleic acids, t structure of which was described by D. L. Woodhouse ar H. S. A. Sherratt in their article on 'The Chemistry of Cells a Chromosomes' in *Science News 24* (pp. 28–34 especially). T statement that a group of enzymes, the dipeptidases, do n necessarily contain protein, is itself of great interest. T generalization that 'enzymes are proteins' has been so wide accepted as to have acquired a quality close to mystic. In th case the contrary conclusion is supported by strong evidenc almost as strong as is possible in support of a negative statemer It raises obvious and disturbing questions for biochemists. Th further statement, that the essential component in the dipe tidases is pentose nucleic acid, is also reasonably well supporte although with reservations which will be discussed. The remai ing statement, that they contribute to protein synthesis, is a obvious hypothesis which has already been put forward and rendered more plausible as well as attractive by Professor Bin ley's work. There is a gap in the evidence between the breakin down type of reaction, used in laboratory work, in which dipeptide is split by the enzyme into its component amino-aci

and energy is released, and the reverse reaction in which energy is absorbed and which, therefore, could proceed only if coupled to an energy-producing reaction. Examples of such coupled reactions are known; and any enzyme, including the dipeptidases, should be able to operate in either direction, if this condition is met. But it is still a hypothesis, even if ideas about protein synthesis as a whole seem now to hang together better than before.

Starting from kidney tissues of pig, Professor Binkley has prepared apparently homogeneous extracts which, from chemical tests, are unlikely to contain more than one per cent of protein; which possess far greater activity than any previous preparation characterized as dipeptidase; which are increased in activity by the removal of protein from the original complex; and which retain their activity after treatments which usually destroy proteins.

The evidence from degree of activity is particularly striking. If the whole activity is assigned to any arbitrary percentage of the final preparation, and a molecular weight is assigned to the active component, then it is possible to calculate the turn-over rate; that is, the number of molecules of dipeptide which are split per second per molecule of active substance. If the whole of the final preparation is active, molecular weights of 10,000 and 100,000 would correspond with turn-over rates of 100,000 and 1,000,000 respectively in conditions of maximum activity. From an admittedly crude estimate of molecular weight, the lower figure would appear to be a minimum, and the higher figure is thought more probable. If the activity is attributed instead to the hypothetical one per cent of protein, which in any case should have been destroyed, then the turn-over rates, already high, have to be multiplied by one hundred. A protein-enzyme of even 10,000 molecular weight would thus have a turn-over rate of 10,000,000; and one of 100,000 molecular weight would have a turn-over rate of 100,000,000. Either figure would establish a new order of activity for any enzyme.

The possibility of a simple inorganic catalyst is excluded by the inactivity of the preparations when confronted with dipeptides containing amino-acids in the so-called unnatural (*D*–) configuration; there is no reason why a catalyst, which itself was simple, should be rendered totally inactive by a change in the geometry of the substance acted upon. There is thus little alternative but to attribute the activity to the main, apparently homogeneous,* component.

Irrespective of the identification of this component, there is a point about specificity which is of interest in principle. Whereas the 'purified' dipeptidase is of high activity, preparations which contain a protein component, though less active, may be more specific. The most natural interpretation is that the catalytic property, as such, is associated with the non-protein component, but that the substances on which it can be exercised are affected by the protein. Because the term 'enzyme' suggests protein, Professor Binkley prefers the non-committal description 'catalyst'.

The main component – the non-protein catalyst – corresponds reasonably both in ultra-violet absorption and also in composition in terms of nitrogen, pentose, and phosphorus, with the figures recorded for pentose nucleic acids. It is also supporting evidence (see the article by *Woodhouse and Sherratt* already quoted) that fractions enriched in purines and pentose respectively have been obtained from it. But there is obviously more work to be done, and certain reservations have been reported. One is that normal, commercial preparations of pentose nucleic acid are without appreciable activity. On the other hand, they act as inhibitors of the active material, which could be taken as an indication of resemblance; and fully active preparations have lately been obtained, starting from commercial preparations. A further discrepancy is that the enzyme ribonuclease, which breaks down normal preparations, has no apparent effect on the high-activity preparations either chemically or in reducing their activity. There are other, possibly corresponding, differences in conditions of preparation and in physical properties; acid conditions during preparation were avoided, and the resulting, active samples are more highly viscous. Nucleic acids are, in any

*Preparations are described as electrophoretically homogeneous; chemical homogeneity in the strict sense is at least a dubious concept, when applied to substances of this order of complexity.

case, complex, and are not of unique composition. Professor Binkley's interpretation is that normal commercial preparations have been partly broken down during their separation, and that his own correspond approximately with 'native' nucleic acid.

If the identification is accepted, the observation that preparations containing a protein component show a lesser, but possibly more specific, activity fits immediately into the picture; since the nucleic acids occur commonly as nucleoproteins – as, indeed, has been one reason from the first for taking an interest in them. It is relevant, also, to quote the following paragraph from Woodhouse and Sherratt, as a summary written independently of the present evidence:

> The metabolic functions of the nucleic acids are far from proved, but there is a considerable amount of circumstantial evidence for the assumption that they are involved in protein synthesis. For example, there is always a high concentration of pentose nucleic acid in tissues actually producing protein, such as rapidly growing tissues in embryos and young animals, regenerating liver in the adult after injury or partial removal, as well as in glands such as the pancreas which produces large amounts of protein-containing secretion. Gastric mucosa cells which make the protein enzyme, pepsin, contain much pentose nucleic acid, whilst neighbouring cells which secrete only hydrochloric acid have very little.

Added together, there is reasonably direct evidence, but with some reservations, that pentose nucleic acid acts as a catalyst in the splitting of dipeptides; a presumption or hypothesis that the same activity in reverse contributes to protein synthesis; and circumstantial evidence connecting the nucleic acids with protein synthesis, irrespective of mechanism. There is also some general support for the enzyme hypothesis of protein synthesis from the removal of an objection to it which has been commonly advanced; namely, that there are so many proteins, individually complex, that the number of specific enzymes required for their making would be impossibly big. Hence the alternative picture of the building up of proteins by a sort of moulding procedure, against 'templates' already formed. Professor Binkley's interpretation makes use of both these approaches; his catalysts, being more general in their activity, can be fewer in number, and the modification of their activity by a protein component is a variant of the 'template' picture.

Since plant viruses are, chemically, nucleoproteins, it was a natural next step that Professor Binkley should examine them for enzyme activity. He has done so in collaboration with Dr C. A. Knight of the Virus Laboratory, University of California. All plant virus preparations so far tested have been found to show enzyme activity of the same kind as pentose nucleic acid, and there are indications also of a further activity which could lead to a lengthening of protein chains: the type of activity represented by the equation, $A.B+C.D \rightarrow A.B.C+D$, in which the four letters represent amino-acid components. There are differences between the activities of preparations regarded as enzymes and as viruses. Whereas it is common experience that treatments which separate the protein and nucleic acid components of a virus destroy its activity as such, activity as an enzyme is increased by treatments which break down the protein component. The experiments link well in this respect with the conception of a plant-virus infection as a disorder of the protein metabolism of the plant: activity in protein synthesis is attributed to the nucleic acid component; the self-perpetuation of an abnormal nucleoprotein requires also that the protein component should be present.

In relation to enzymes in general it may be recalled that the plant viruses used in these experiments can be obtained as crystalline material, and that their activity as enzymes is increased by partial breaking down. But it is the exception rather than the rule for an enzyme to have been obtained in crystalline form, nor is crystallinity more than a very rough index of purity with such complex substances. There is provocation for thought and experiment in this juxtaposition.

On the wider aspects of protein synthesis, 'it is possible by a little thought' – quoting now from Professor Binkley – 'to arrive at a broad framework of a mechanism of synthesis of specific proteins and the genetic control of that synthesis. However, in a broader sense little has been done except to shove the problem

back one step. Before an answer is possible we must consider how a nucleic acid may behave as an enzyme and, at the same time, direct the synthesis of itself.' Of nucleic acid, that is, as well as of protein.

CLOUDS AND RAIN-MAKING

Research into the physics of clouds, rain, and rain-making has been proceeding for more than five years in Australia on a scale which is readily justified in relation to conditions. It is being carried out by the Divisions of Physics and Radiophysics of the Commonwealth Scientific and Industrial Research Organization, and is reviewed in the report of that body for 1951–2. One of the results of this programme has been the demonstration of rain-formation in clouds warmer than freezing point, apparently through the coalescence of droplets of different sizes when brought into collision by up-draughts. A survey of rain-storms, made in and around Sydney during a twelve-month period, largely by radar methods, has established that some 15 per cent of rain in this area is from clouds in which only water droplets are present.

There have been other quantitative surveys: one of the size-distribution of droplets in clouds, measured by the impact-holes left on slides coated with magnesium oxide and exposed from aircraft; and another to obtain estimates of the total quantity of liquid water in convective clouds. The quantity at all levels above the base of a convective cloud has been found to be considerably less, under Australian conditions, than that suggested by theoretical studies. The suggestion is being investigated that the discrepancy is caused by the rising from ground-level of air dry enough, even after expansion and cooling, to bring about the evaporation of droplets already formed. Strips of light metal foil are being used to study air-flow patterns around cumulus clouds; the foil being released from an aircraft, and its vertical movement followed by radar from the ground.

The frequency of cloud-conditions suited to 'seeding' with solid carbon dioxide – as used in Australia in February 1947 – have been assessed, also in the Sydney area. Marks were allotted

on a system based on synoptic data, height of cloud-top in relation to freezing level, and water-content above freezing level. Within a few hundred miles of Sydney, it was concluded that opportunities for successful rain-making by this method would average less than one day per week over the whole area. The same statistics should apply broadly to seeding with silver iodide released from the air. The release of silver iodide from smoke generators on the ground, for which spectacular claims have been made in the United States, is discussed critically in the report. For three months in 1951, a smoke generator was operated at Hay, in the dry Riverina district of New South Wales, and an analysis of the records by rain-gauge observers within one hundred miles is still in progress. Rainfall in the sector down-wind from the point of release was greater than up-wind; but preliminary results suggest that the increase was not great enough to distinguish between silver iodide and chance.

Moderate successes are reported for a method in which rain is produced by spraying water into the base of a suitable cloud. This was an Australian idea, having arisen out of studies of rain-formation by coalescence. It has led in turn to a further method which is still at an early stage of experiment. The idea is to disperse large electrically charged nuclei, also from an aircraft. Like the water-spray method this would be independent of cloud-top temperature, and it is hoped will prove a more economical procedure than others tried. The connexion is emphasized between the fundamental researches which are in progress and attempts at application.

In relation to the discussion of 'effective rainfall' given by R. N. Elston in his article on 'Desert Research' published in *Science News 27* (pp. 22–24 especially), it is of interest that the Section of Meteorological Physics at Edithville, Melbourne, has developed a method of measuring the rate of evaporation from soil or other natural surfaces which is described as, in principle, the only one which can be applied to any type of surface. An ultimate aim of this work is to assess the net loss of water by evaporation from any region at any time. In present usage, the allowance for evaporation in 'effective rainfall' is based on the

rate of evaporation from a free-water surface. The problem of heat balance is being studied concurrently.

MICROSCOPY OF FINE PARTICLES

A new technique for the examination of fine particles of materials, such as coal, which under normal conditions are opaque, but become translucent if sufficiently thin, was shown by Dr R. G. H. B. Boddy of the Medical Research Council Pneumoconiosis Research Unit at a one-day symposium and exhibition on microscopical methods held lately at King's College, London, by the Industrial Section of the Royal Microscopical Society in collaboration with other interested bodies. What is unusual in the method is the introduction of a 'crushing condenser' so that the squeezing out of the particles between specimen-slide and condenser lens can be directly observed. Unconventional as it is, the method has been found useful in practice, and should be capable of other applications. For obvious reasons, it is necessary that the larger particles should first have been removed by effective gravity settling.

CORRESPONDENCE

BALLOONING FOR RESEARCH

MANY must have read with interest the excellent article by Herz and Tennent, in *Science News 27*, on 'Ballooning for Research'. In one respect, however, the authors are perhaps too dogmatic, and some comment is called for. It is stated that 'the disadvantages of the closed balloon lie in its inability to float at altitude', and Figure 3 in the article shows the flight curve of such balloons in contrast to the ideal flight curve of Figure 1. Experience in India, however, has shown that a cluster of closed balloons does find a definite floating level, in which it often remains for many hours. In some flights there may be three or four such levels.

These features of balloon behaviour were very clearly brought out in the course of cosmic-ray experiments in South India conducted by the Tata Institute of Fundamental Research towards the end of 1950. The flights were made with neoprene balloons. The altitude data were obtained from a series of trigonometrical stations spaced across a wide stretch of country, and most of the flight curves are known with accuracy. In one flight a cluster of balloons maintained a steady height of 60,000 feet for nine hours, with very small fluctuations. The behaviour, in fact, reproduced almost exactly the ideal curve of Herz and Tennent's Figure 1.

To explain these results, it seems very probable that the assumption 'that the skin of the balloon has no influence apart from separating the gas inside the balloon from the air surrounding it' is not wholly true. At the low pressure of the upper atmosphere the tension in the skin is no longer a negligible factor. A balloon which is perfectly extensible would not float, but one which is not so begins to approximate in its behaviour to the constant-volume balloon, and like the latter has a floating level. A cluster of such balloons may assume one of several levels,

depending on the number which survive. A fuller account of this has been published.*

Rubber balloons deteriorate rapidly under the action of ozone and ultra-violet light, and prolonged floating is therefore unlikely. With neoprene balloons there seems to be a real difference between the behaviour in the tropics and in northerly latitudes. Neoprene becomes less extensible at low temperatures, as Herz and Tennent point out, and any differences of balloon behaviour in different latitudes is almost certainly a temperature effect. The stratosphere temperature in the tropics is markedly lower than in high latitudes, and it may be for this reason that the floating behaviour which is such a striking feature of the Indian flights has not been encountered elsewhere.

H. J. TAYLOR

Department of Physics, Wilson College, Bombay

**Proceedings of the Indian Academy of Sciences*, A 36, 41–54 (1952).

Dr A. J. Herz and Dr R. M. Tennent write :

We are very grateful to Professor Taylor for his comments which allow us to say a little more about the flying of expansible balloons. We are, of course, aware that clusters of balloons have been flown by a number of workers, including those at Bombay, and that successful flights have sometimes been obtained. However, the open-necked constant-volume balloons are far more reliable, and their floating altitude can be predicted and adjusted *before* the flight. Dr Taylor's own results (*loc. cit.*) show that the behaviour of clusters tends to be rather erratic; a really level flight was obtained only on one occasion. In practice also, their floating altitudes seem to be on the low side: 60,000 feet, for example, is not a satisfactory altitude for most cosmic-ray investigations.

We entirely agree that the tension in the balloon fabric is a quantity which cannot be neglected in accurate calculations, and we have, in fact, taken it into account in our work. The magnitude of this tension, however, depends on the physical conditions

in the atmosphere, and on the treatment the balloon underwent before it was launched. To this must be added the complication that a certain amount of plastic flow occurs in stretched neoprene, so that the tension will depend on time also. As a result it is difficult to take account of the tension in a way which allows one to predict the balloon performance.

Professor Taylor's suggestion on the effect of the low stratosphere temperatures encountered in the tropics is a very reasonable one, though more evidence is still needed. Attempts have indeed been made by several workers, we understand, to exploit this effect by launching balloons about 1½ hours before sunset. During the ascent the balloons are kept warm by the sun's radiation. After sunset they stiffen, and the increased tension may result in a stable floating level being maintained. Professor von Friesen's group in Sweden have obtained one successful level flight at a very great altitude by this method, but it is too early yet to appraise it fully.

THE VELOCITY OF ELECTROMAGNETIC WAVES

We are indebted to a correspondent at the University College of North Staffordshire for pointing out that the probable error associated with the determination by Smith, Franklin, and Whiting, mentioned in Dr W. C. Vaughan's article in *Science News* 27, was ±50 km. per second (as correctly indicated in Fig. 1, p. 66), *not* ±5 km. per second, as given on p. 64.

SOME BOOKS RECEIVED

THE COMETS AND THEIR ORIGIN by R. A. Lyttleton (Cambridge University Press, 1953), pp. 173, 25s.

A compact summary of observational knowledge about comets, illustrated with thirteen plates, is given as an introduction to the second and longer section of the book, in which Dr Lyttleton presents in some detail a theory of the origin of comets by 'the accretion of interstellar dust through the gravitational action of the sun during passages through interstellar dust clouds'. Described as 'one of the principal successes of the New Cosmology', theory and book can as well be considered on their own.

EVOLUTION IN ACTION by Julian Huxley (Chatto & Windus, 1953), pp. 160, 9s 6d.

Dr Huxley's 'map of the evolutionary process' is widely enough planned to include the development of mental activity, the organization of cumulative knowledge, human values and, in part, society; the whole being related to his general proposition that biological progress is 'improvement which permits or facilitates further improvement'. The book is based on the author's Patten Foundation Lectures, given at Indiana University in 1951 and later adapted for broadcast.

THE BILE PIGMENTS by C. H. Gray (Methuen, Monographs on Biochemical Subjects, 1953), pp. 142, 9s 6d.

The enormous, and largely recent, literature on the bile pigments is the best illustration of the interest which has developed from what in the first place was the purely medical problem of jaundice. The present compressed, but mostly readable, summary should be of value to a mixed audience of biologists, chemists, and biochemists working in other fields, and medical students.

PHYSIOLOGY OF VISION (*British Medical Bulletin*, Vol. 9, No. 1: London, British Council, 1953), pp. 84, 15s.

A symposium by leading British workers in this field, edited by W. S. Stiles, with bibliography and reviews.

ABOUT OUR CONTRIBUTORS

J. L. CLOUDSLEY-THOMPSON is a lecturer in zoology at King's College, London. Born in Kashmir in 1921, he served as a captain in the 4th C.L.Y. 'Sharpshooters' during the war. Returning to Cambridge in 1945, he wrote a Ph.D. thesis on the physiology and ecology of millipedes, and took part in an expedition to N.W. Iceland. He is honorary secretary of the Biological Council.

J. CROWTHER graduated in metallurgy at Manchester, and did research at Sheffield; began his industrial career in steel, but later became connected with the ancient metal, copper, and the newcomers, aluminium and magnesium. Sometime Research Metallurgist, and now Chief Metallurgist of James Booth and Co., Birmingham; likes to temper his metallurgy with a little physics.

R. O. DAVIES, a 29-year-old New Zealander, graduated at Otago in mathematics and physics. He went to Oxford as a Rhodes Scholar in 1948, and acquired under Professor Simon a taste for thermodynamics and physical statistics. He is continuing research in these subjects at Queen Mary College as an I.C.I. Research Fellow of the University of London.

M. F. DAY was born in Sydney, New South Wales. He studied entomology at the University of Sydney, and at Harvard University. Has worked mainly on insect physiology, and is now studying mechanisms of virus transmission by insects at the Division of Entomology, Commonwealth Scientific and Industrial Research Organization, Canberra, Australia.

FRANK FENNER was born in Victoria, and educated at the University of Adelaide. After six years as a medical officer in the Australian Army, including three years as a malariologist in New Guinea and Borneo, he worked on animal viruses under Sir Macfarlane Burnet, in Melbourne. Spent a year at the Rockefeller Institute in New York in 1948–9, and was then appointed Professor of Microbiology in the Australian National University, Canberra.

PAUL V. SMITH is aged 31, received a Ph.D. in organic chemistry from the University of Illinois, and, after a year of post-doctoral work there, joined the Standard Oil Development Co. in 1945. Now

working on various aspects of geochemistry; author of s papers and patents in the fields of synthetic rubber, synthet cants, chemical separation processes, and geochemistry.

S. VAIDA studied mathematics and physics and took the deg Dr.phil. at the University of Vienna. He worked as an actua various countries, and came to England in 1939. In 1944 he jo the staff of the Admiralty, and is a member of the Royal N Scientific Service.

(2)	(1)	(2)

SCIENCE NEWS

Outstanding articles from recent issues include :

SCIENCE NEWS 24

The Measurement of Mental Images: *P. L. Short*
Chemistry of Cells and Chromosomes: *D. L. Woodhouse and H. S. A. Sherratt*
Double Stars: *W. H. Marshall*
The Periodic System: *H. R. Paneth*
Hydroelectric and Irrigation Development in the U.S.S.R.
Statistics as an Aid to Literary Studies: *C. B. Williams*

SCIENCE NEWS 25

Concepts out of Context: The Pied Pipers of Science: *N. W. Pirie*
Radio-Frequency Methods in the Study of Atoms and Molecules: *J. Little*
Dietary Self-Selection Experiments: *D. E. Tribe*
The Mechanism of Mineral Salts Absorption by Plants: *J. F. Sutcliffe*
Colour Vision: *Katharine Tansley and R. A. Weale*
Current Lines of Research and Development in Gas Turbines: *J. Hodge*
Under-Water Television

SCIENCE NEWS 26

The Norwegian–British–Swedish Antarctic Expedition 1949–52: *E. F. Roots*
A New Era in Microscopy: *J. G. Thomas*
Ultra-Violet Radiation in the Examination of Works of Art: *George Savage*
Mechanisms of Chemical Change: *K. D. Wadsworth*
Ion Exchange: *K. S. Spiegler*
Titanium: A New Metal for Engineers: *F. A. Fox*
Philosophical Analysis and Scientific Progress: *John E. Owen*

SCIENCE NEWS 27

Desert Research: *R. N. Elston*
Cosmic Radiation: *E. P. George*
Ballooning for Research: *A. J. Herz and R. M. Tennent*
The Velocity of Electromagnetic Waves: *W. C. Vaughan*
Social Scientists in Action: *Anthony H. Richmond*
The Szilard-Chalmers Effect and the Separation of Radioactive Isotopes: *D. G. Tuck*