24 9143
E4

The Measurement of Mental Images
Chemistry of Cells and Chromosomes
Double Stars
The Periodic System
Hydroelectric and Irrigation Development in the U.S.S.R.
Statistics as an Aid to Literary Studies
Taking a Sample
Research Report

THE MEASUREMENT OF MENTAL IMAGES

P. L. SHORT

THINKING about thought is difficult. Self-reflexion or introspection is needed, which many modern psychologists want to ignore in their studies of human behaviour. Yet we can at any rate investigate the *materials* of thought, namely, mental images. With mental tasks or problems, what matters psychologically is not so much the answer arrived at, but the kind of imaging experienced. Some method of registering what is happening physiologically when imaging occurs is essential to complete the story. In brief, a psycho-physiological account of imaging is the goal.

Consider the following puzzle. The best way is to read it through, then close your eyes and work it out. 'Think of a solid cube, painted red on the outside only. Cut it in halves, then in quarters. How many sides are painted red and how many are unpainted?'

Images will emerge in your mind as you reflect; probably a picture of a cube, or you may find yourself merely counting the sides. If you try it on your friends, you will find differences in the ways they tackle the problem. Certainly the solution will come more readily to those who can picture the cube in their minds, especially if they can see the cube in varying positions, than to others who have to count the sides or work it out mathematically. In short, two chief modes of imagery emerge, visual and verbal.

There is nothing novel in merely classifying imagery. Imagery differences were first observed experimentally by Francis Galton in 1880. He sent questions to his acquaintances asking them to visualize their breakfast tables, noting differences in the

vividness of the visual images that they obtained. Other kinds of imagery have since been studied, chiefly those of sound (auditory or verbal images), and touch (tactile or kinaesthetic). Psychologists have assessed the relative predominance of these kinds of images in different people by using rating scales. The subject is scored, or scores himself, according to the vividness or clarity of the images that he has when asked to imagine himself seeing or hearing different items. If asked to picture the front of his house, he may either visualize it clearly (scoring 3), have a moderate visual image (2), or experience only the vaguest mental glimpse (1). This kind of test does not permit the use of objective standards with which we can compare one person's reports with another's. Above all it is highly artificial.

One is often perplexed to know how to study the mental life of the individual in the laboratory whilst using the necessary scientific controls. One wants to avoid imposing on the subject too strained and bizarre an intellectual régime. In the study of thinking, the vividness or clarity of single memory-images is really only of academic interest to imagery classifiers. What needs to be shown instead is the way we employ images in familiar mental tasks within a controlled situation, securing at the same time a record of relevant physiological events. We can then compare or correlate what the subject reports *after* task performance with physiological recordings taken *during* the performance itself. This is one way of checking introspections, at any rate those parts of them in which we are interested.

In discussing the 'Cube Problem' we noticed two main types of imagery, visual and verbal. Let us examine, first, the visual image. An experiment by C. W. Perky in 1910 showed that there is no clear division between actual seeing and visualizing. He asked subjects to project on a screen a visual image of a banana. At the same time he threw a faint picture of a banana on to the screen, increasing its intensity until subjects said they had a good image. Now the image fluctuated as he varied the intensity of the displayed picture. This striking experiment, together with similar evidence, enables us to say that actual sensations and images may be confused, and are

therefore in a similar psychological category. We are encouraged to look for a common physiological mechanism to account for both.

Up to the present the main kind of body-measurement associated with mental features has been the measurement of physique. Work on these lines was begun by E. Kretschmer, a psychiatrist, who classified the characteristic body-builds of certain types of mental patients. Recently, W. H. Sheldon, working with normal people, chose three aspects of physique as his basic variables. He then presented 4,000 photographs of young male physiques to trained judges and asked them to arrange the photographs in rank-order for each variable, on a 7-point rating scale. A glance at this method is enough to show that we could never employ any measurements like this for recording mental tasks.

G. M. Wyburn in *Science News* 23 has described the physical chain of events that occurs when we see an object; it culminates in electrical activity in the posterior parts of the brain. This activity is not great enough to be observed directly through the skull, although an indication of it can be obtained in many individuals from the suppression of electrical rhythms which are otherwise present. These rhythms are synchronized in many cells, and so can be detected through the skull.

The recording of electrical rhythms from the brain by means of the electro-encephalogram (EEG) was described by W. A. Cobb in *Science News* 14. The best-known rhythms from the posterior parts of the brain are the alpha-rhythms, with a rhythmic fluctuation in voltage at a frequency of 8-13 c./s. Cobb mentioned the individual variations that can be observed in the amount of alpha-activity. It was firmly established in 1943 that alpha-rhythms are recorded from many people only when they close their eyes – the *R* (Responsive) type. With others the rhythms are continually present, whether the eyes are opened or closed – the *P* (Persistent) type. In the third group, no alpha-activity is ever recorded – the *M* (Minus) type. We will now consider the possible relation between alpha-types and imagery-types.

It has been shown in EEG recordings that when a subject of the *R* type works out a simple problem in mental arithmetic with his eyes closed, his alpha-rhythms sometimes diminish in amplitude, or 'block' entirely, in much the same way as they do when his eyes are open. When this happens he very often reports afterwards that he has been visualizing the numbers. An example of a record obtained in this way is shown in Fig. 1.

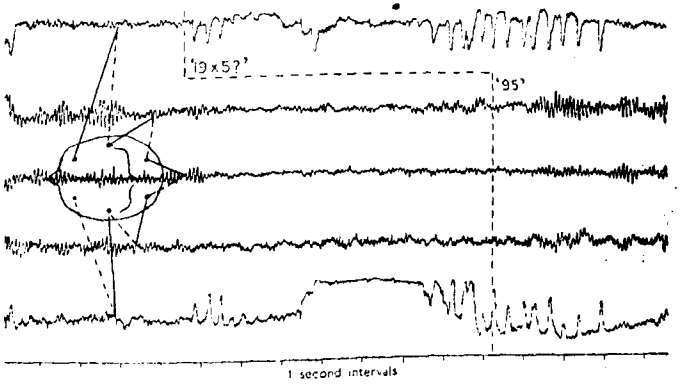


Fig. 1. EEG of a subject with a 'Responsive' alpha-rhythm. The record shows a characteristic blocking of the alpha-rhythm when the sum is given. The rhythm returns to its previous amplitude once an answer is found. The diagrammatic head shows the positions of the electrodes: each tracing is of the difference in potential between two electrodes, shown by pairs of lines connecting the electrode-positions to the tracing. Vertical scale: 0.1 inch = about 25 microvolt.

This gives us a clue to the way visual images in our thinking processes can be objectively studied. But what kind of imagery occurs when the alpha-rhythms persist unchanged during thinking? This brings us to the second, verbal type of image. When subjects fail to visualize numbers, then they tend either to say or to count them.

In 1929, F. L. Golla and S. Antonovitch found amongst the people they examined two main types of breathing or respiratory movement, regular and irregular breathing. They found, in addition, that the regular respiratory type used mainly visual images in thinking, and the irregular type, verbal images. Irregular respiration was believed to be due to the tendency to 'voice' words or numbers, occurring together with the movement of the larynx.

Up to this stage, we can say that EEG method gives indications which may correspond positively with the use of visual imagery, and the respiratory method indications which may correspond positively with the use of verbal imagery. If we combine the findings from the two methods, we can make a comprehensive study of the two main types of imagery. Accordingly, simultaneous recordings of EEG alpha-rhythms and breathing have been made recently with 150 subjects under strictly experimental conditions. The subjects were drawn in equal numbers from two groups: 75 were out-patients, mainly epileptic, coming to the Burden Institute at Bristol; and 75 were individual volunteers, mainly university students. In the out-patient group the age-range was 8–59 years. Only those patients with no marked abnormalities in the electrical activity of the parieto-occipital areas of the brain, and with no obvious psychological disorientation, were selected. In the student group the age range was 18–37 years. The intention was that subjects in the latter group should act as experimental 'controls'. A careful analysis, however, failed to reveal any significant differences between results obtained from patients and students. They are therefore treated as a single group in the results.

METHODS AND PROCEDURE

A preliminary examination on a 6-channel EEG enabled subjects to adjust themselves to the conditions of the experiment. The main examination was given with a two-channel portable EEG machine; the first channel recorded electrical potentials from the posterior areas of the brain, and the second channel recorded breathing by means of a thermocouple. The latter was used to give a record of temperature variations in front of the face (cooler during breathing-in, warmer during breathing-out).

It was essential for the subject not to know the real purpose of the apparatus, so it was fixed to a spectacle-frame. None of them guessed what it was for. Had they done so, their breathing would have been affected in unpredictable ways. To decide whether a subject's breathing was regular or not, each breath had to be measured. The distance from one inspiration to the next was measured in centimetres on the record, and converted into seconds throughout. Each breath was then plotted on a 'breath time chart'.

The first part of the main recording was devoted to the discovery of alpha-type and habitual mode of breathing while the subject was at rest. Six mental tasks were then given in this order, with standard inquiries: (1) *Sum.* 'Will you multiply 19 by 5?' (2) *Story.* 'Think over the story of Noah's Ark.' (3) *Cube Problem.* As given above. (4) *Prayer.* 'Think over the Lord's Prayer.' (5) *Anthem.* 'Think over the National Anthem.' (6) *Argument.* 'Think over points for and against the Labour Party, and for and against the Conservative Party.'

After completing each task, the subject was asked: 'In what way did you think about that? What was going on in your mind when you were thinking of (for example) the National Anthem?' Replies were written as given directly on the records, and later carefully classified.

In assessing the breathing records, two factors having nothing to do with imagery, but which occasionally caused irregular breathing, were discounted from the results. The first factor was an occasional restless anxiety or fidgeting, recorded on the EEG as muscle action-potentials in the form of a fast 'spikey' activity.* This response was shown by five subjects; they were considered unsuitable for experiment, and were replaced by others. Secondly, at the outset of the more difficult tasks like the 'argument', a certain tension occurred in some ten of the subjects: breath was held or inhibited for up to 14 seconds on the records, and alpha rhythms were blocked. The respiration-effect was shown in similar recordings by Golla and Antonovitch in 1929. The occurrence of both effects

* See Floyd and Silver, *Science News* 22, for similar recordings.

together is the justification for ignoring the beginning part of such records in relation to the type of imagery used. What happens is that the individual is summoning all his resources to deal with an emergency situation; there is a mild intellectual crisis. Once the material is mustered, the tension disappears quite quickly. When this happens, images emerge and are utilized, as indeed the images *are* the thought-material.

RESULTS

Breathing Types and Imagery Types

Extracts from the 'breath time charts' for three individuals are shown in Fig. 2. Times between breaths when resting and when doing tasks are distinguished on the chart. With subject (1) the difference between the shortest breath (2.3 sec.) and the longest (6.6 sec.) when resting is 4.3 seconds. He is an irregular breather. So also is subject (2). But with subject (3) the span is only 0.5 sec., and he is a regular breather. Also from Fig. 2, it can be seen that the shorter the span when resting, the greater is the regularity of breathing when doing tasks, and *vice versa.* This is a general rule.

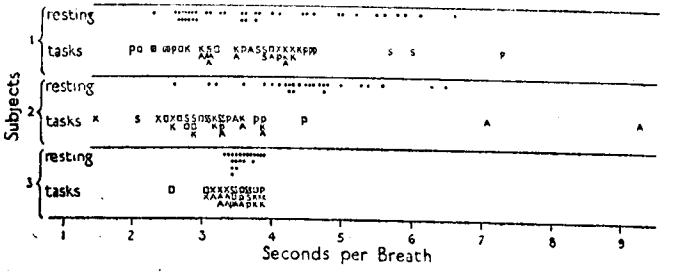


Fig. 2. Extracts from 'breath time charts' of three subjects, resting and during task periods. Subjects (1) and (2), irregular, resting and during tasks; subject (3), regular. Breaths recorded during resting are shown • in the upper half of each chart; breaths during tasks, in the lower half. Breaths during the *Sum* tasks are shown, X; *Story* task, S; *Cube Problem*, □; *Prayer* task, P; *Anthem* task, K; and *Argument* task, A.

Of the subjects examined, 76, or 50.7 per cent, were regular breathers; and 74, or 49.3 per cent, were irregular.

When the imagery classifications were examined, it was found that the subjects fell into two groups. Either they reported that they were aware of having mainly visual images when doing tasks, or mainly verbal images. The former were classed as 'Visualists', the latter as 'Verbalists'. Of the 80 Visualists, 66, or 82.5 per cent, were habitually regular breathers; only 14 of the Visualists, or 17.5 per cent, breathed irregularly. On the other hand, of the 70 Verbalists, 60, or about 86 per cent, breathed irregularly all the time; and only 10 Verbalists, or about 14 per cent, were regular breathers. These results are displayed in Fig. 3, which shows at a glance the general correlation between breathing spans during resting, and subjective assessments.

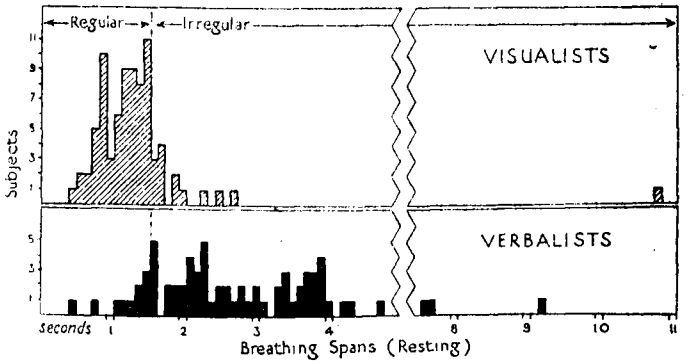


Fig. 3. Histogram plotting regularity and irregularity of breathing for all subjects, classed as Visualist or Verbalist from subjective reports. Horizontal baseline shows the breathing spans (see text) from 0.5 sec. to 10.8 sec. The number of subjects for each span (in steps of 0.1 sec.) is plotted vertically.

During task-periods, the habitual regularity or irregularity of both groups tended to increase. Primary visual or verbal imagery, respectively, was reported. This supports the finding

of C. Fox in 1914 that images occur in thinking chiefly when a mental difficulty or problem arises. However, provided that the subject is not actually asleep, he will presumably be thinking about something; that is why regular or irregular breathing is shown even when no set problem is being tackled.

The figures given above of associations between breathing when resting, and type of imagery, were submitted to a standard statistical analysis*. The odds that the associations obtained between imagery and customary breathing are due to chance, are more than 10,000 to 1 against – a very safe indication of a highly significant relation between them. Similarly reliable statistical results were obtained with all tasks.

Alpha-types and Imagery-types

Let us consider next the alpha-types. Typical records are shown in Figs. 4–6. It was found that the great majority both of Visualists (56) and of Verbalists (43) had records of the *R* type (Fig. 4), responsive by suppression of the alpha-rhythms. There remained 18 Visualists, and only 3 Verbalists, who were *M* type (Fig. 5), with no alpha-rhythms; and 24 Verbalists, and only 7 Visualists, who were *P* type (Fig. 6), with alpha-rhythms persistent. Imagery with these last two groups tended to be of an extreme nature; exclusively visual with the *M*'s and almost constantly verbal with the *P*'s. The *R* group was the largest, and was intermediate between the other two, showing mixed imagery, but with a definite predominance of visual or verbal images in individual subjects. The fact that the two distributions of alpha-type and imagery-type significantly overlap each other, is shown by the same statistical method. The odds against the associations between these two types, alpha and imagery, being due to chance are again more than 10,000 to 1.

[Figs. 4–6, with explanatory notes, are on the two pages following.]

*The χ^2 method was used. χ^2 was calculated as 40.62, and the appropriate 'degree of probability', quoted in the text, found from a χ^2 table.

Fig. 4.

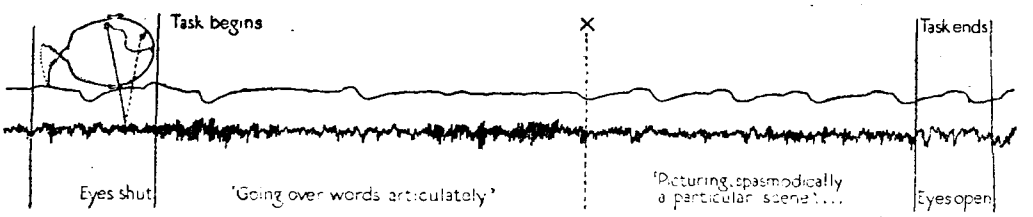


Fig. 5.

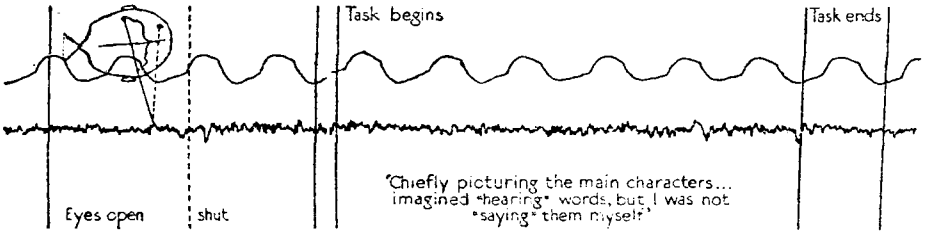
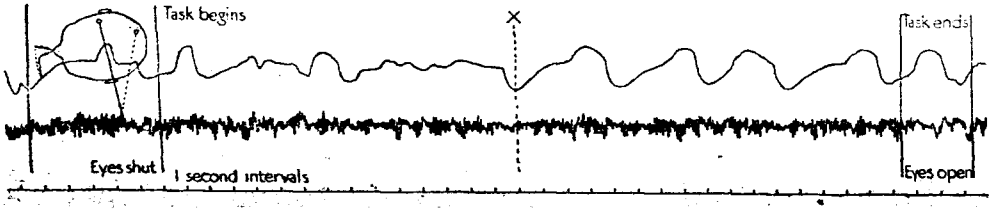


Fig. 6.



Specimen records obtained from three subjects with a two-channel EEG machine. The upper tracing from each subject is of breathing, recorded with a thermocouple (see text). The lower tracing is of electrical potentials from the posterior parts of the brain. Electrode-positions are shown diagrammatically as in Fig. 1. The time-scale, shown at the bottom of Fig. 6, is more compressed in these combined records than in EEG records taken alone (c.f. Fig. 1). Vertical scale: 0.1 inch = about 30 microvolt. Records are divided into three sections: resting with eyes shut; doing 'task'; and eyes open. Alpha types are assessed by comparing the EEG records when resting with eyes shut and with eyes open respectively. In the descriptions of individual records, 'span' means the difference between the longest and shortest intervals between in-breaths. The beginning and end of 'tasks' is marked on the records.

Fig. 4. Alpha-type: *R*. A 'Verbalist'. During resting, respiration is irregular (span, 3.5 seconds). The 'Argument' task-period is illustrated. Respiration is irregular up to 'X', then regular (span, 0.6 second). Note corresponding behaviour of alpha-rhythms.

Fig. 5. Alpha-type: *M*. A 'Visualist'. During resting, respiration is very regular (span, 0.5 second), and throughout all tasks. The 'Story' task-period is illustrated, and respiration remains completely regular (span, 0.2 second). Note the absence of alpha-rhythm at all times, in contrast with Fig. 6 (*P*-type) in which alpha-rhythms are at no time wholly absent.

Fig. 6. Alpha-type: *P*. A 'Verbalist'. Respiration very irregular during resting (span, 4.3 seconds). The 'Prayer' task-period is illustrated. Respiration is irregular up to 'X' (span, 4.4 seconds), then becomes more regular (span, 1.0 second). Alpha-rhythms persist throughout. Subject evidently verbalizing up to 'X', but not afterwards. No visual imagery. Report, after completion of task: 'I didn't get beyond the first sentence (claimed to be articulated). I was puzzled after that, and didn't "say" any more.'

Task-periods were examined in detail in order to see how often the alpha-rhythms blocked when the subject was visualizing, and how often they persisted when he was verbalizing. The result was quite conclusive. Out of 833 instances of imagery-reports and corresponding alpha-activity, 517, or 62.8 per cent, were alpha-responses with visual imagery, and only 67, or 8 per cent, with verbal imagery; 200, or 23.3 per cent, were alpha-persistences with verbal imagery, and only 49, or 5.9 per cent, with visual. Detailed statistical analyses of each task again simply confirmed the main analysis.

Summary of Results

Results then are clear. The subjects fall into two main categories of imagery, visual and verbal. The extreme alpha-types correlate with extreme types of imagery, and the *R* type with 'moderate' imagery, yet with a clearly recognizable predominance of visual or verbal images. The Visualists tend always to breathe regularly, and their alpha-rhythms, where present, block whenever they are busy with mental tasks; visual images come more readily to them than verbal ones. The Verbalists tend always to breathe irregularly, and their alpha-rhythms tend to persist, whether they are thinking out problems or not.

Other modes of imagery occurred now and then in the reports, but did not predominate. The kinaesthetic (movement) image is of particular interest, and is now being studied by similar methods.

IMAGERY IN MENTAL LIFE

How images actually function in mental life is more a matter of speculation. J. A. V. Butler's approach in *Science News* 22 fits in well with the experiment described. He relates how we create a model of our surroundings in order to plan appropriate behaviour. It is perhaps misleading to speak always of 'pictured' models, when we find that the extreme *P* types create only verbal models, with little or no visualization. Experimentally, it appears that we not only create models of the world

to plan overt actions, but use models as images to deal with the problems of thought.

Some theorists (e.g. the 'Wurzburg School') believe that thought is essentially imageless; they base their views on introspecting findings alone. They sometimes readily admit that images abound in thinking, but claim that a 'consciousness' or 'awareness' (*Bewusstseinslage*), difficult to define, is *really* what characterizes thought as such. May not this 'awareness', however, be simply another way of expressing the manipulation of our image models in thinking?

In the experiment, no special study was made of intelligence, nor of the subject's *ability* to deal with individual tasks. It is arguable that intellectual skill resides only in the manipulation if images, however restricted we may be in the type of images at our disposal. The more an individual can adapt his models to suit his thinking, the greater may be his efficiency in dealing with a variety of mental tasks. But if we can utilize visual or verbal images with about the same ease, then clearly we are better fitted to deal with more problems, whether they demand mainly visual or verbal aids.

Of all the alpha-types, persons belonging to the *R* group are the most favoured in so far as they can readily adapt their models to suit different types of mental problems. While not obsessed with private pictures, they can evoke satisfactory visual patterns when necessary, and can combine data from the various sense organs more readily than can those of either the *M* or *P* types.

The subject, a *P* type, whose record is given in Fig. 6, had his alpha type been different, might have recollected the words of the remainder of the prayer by visual means, instead of trying to recall them by purely verbal methods. But visual imagery was wholly absent from his reports, and presumably from his mental life as a whole.

Similarly the 'argument' demands some use of words. A subject of the *M* type, an extreme Visualist, who never verbalized, shows just how bizarre exclusive visualizing can be with this type of problem.

A map of England with the centre at London, with black lines radiating, symbolizing nationalization. The Conservatives are represented by spots without connecting lines, with the words 'Free Enterprise' written. I was neither hearing nor saying anything.

From the evidence, statistical and experimental, now available, it seems that alpha frequencies are inborn and probably hereditary. To some extent this is true of the alpha-types, chiefly of the *M*'s and *P*'s. In 'identical' twins the resemblance between alpha frequencies is as great as that between fingerprints; but there are differences in the details of the response to stimulation. For instance, alpha-rhythms normally responsive only to light can become conditioned to sound instead, when the sound is repeatedly presented with the light. This gives another instance of the adaptability of the *R* type. The type of imagery experienced by the individual is not necessarily innately determined.

So far, the argument has been in favour of those of the *R* type rather than the others, on the grounds of superior imagery-adjustment. What are the intellectual chances of the minority groups *M* and *P*, in a world of infinite opportunities? Given congenial living conditions, an *M* type may persevere in dealing with the kind of problems which by nature he is eminently suited to deal with. The same holds for the *P* type. Both may stand better chances of becoming geniuses than the adaptable *R*, whose infinite capacity for adapting himself may turn into mediocrity; he becomes brilliant at nothing in particular.

The models employed in imagery are always with us as long as we are capable of consciously reacting to our surroundings; the habitual modes of imagery with their particular physiological characteristics have shown us so much. If this is so, then they play an important part in the matrix of total psychological adjustment. There is only a short step between the measurement of imagery and the measurement of personality. In future, the kind of objective recording methods we have described may well be used to throw new light on the toughest problems of the human mind.

ACKNOWLEDGEMENT

I should like to thank Professor F. L. Golla for permitting publication of results.

FURTHER READING

Galton, F. *Inquiries into Human Faculty*. London, Macmillan, 1883.
Woodworth, R. S. *Experimental Psychology* (Ch. 2). New York, Henry Holt. (For the *Perky* experiment.)
Kretschmer, E. *Physique and Character*. London, Kegan Paul, 1925.
Sheldon, W. H. *The Varieties of Physique*. New York, Harper, 1940.
Golla, F. L., Hutton, E. L., and Walter, W. Grey. *J. Ment. Sci.*, **89**, 375 (1943).
Golla, F. L., and Antonovitch, S. *Brain*, **52**, 491 (1929).
Golla, F. L., and Antonovitch, S. *J. Ment. Sci.*, **75**, 234 (1929).
Fox, C. *Brit. J. Psychol.*, **6**, 420 (1914).
Humphrey, G. *Thinking*. London, Methuen, 1951. (For the *Wurzburg* School.)

THE CHEMISTRY OF CELLS AND CHROMOSOMES

D. L. WOODHOUSE AND H. S. A. SHERRATT

THE cell is a self-maintaining system with the chemical and physical mechanisms for obtaining material from its environment to satisfy nutritional and energy requirements. The fertilized ovum is so constituted that it can develop into a differentiated organism of the same type as its parent. It is not surprising, therefore, that cells are complex in structure and composition. This has been appreciated in some degree for many years, although much of the earlier work on the biochemistry of tissues dealt of necessity with an organism or organ as a unit and was concerned chiefly with the chemical changes in a tissue or with its chemical composition as a whole. There was therefore a gap between fine structure, which could be observed, and fine composition, which could not yet be studied. More lately, a number of new techniques have been successfully applied in isolating and examining the chief components of the cell.* Some of the researches in which they have been employed have already opened a new chapter in the biochemistry of living tissues: one relating the chemical composition and the properties of a number of sub-units which chemical and biological examinations have shown to exist in the cell. It has been possible to collect some of these sub-units in a relatively pure form and in sufficient quantity for chemical analysis. A phase has thus been reached when studies can be undertaken on the cells themselves or on their sub-units.

Cells are not static structures and there is an intimate relation between structure and function of a mutually determining character. We have come to realize that the different chemical reactions occur in an organized manner, governed in part by the structure of the cell. By getting more exact knowledge of the composition of the sub-units we can more fully appreciate the spatial co-ordination of these reactions.

*See, for example: Harris, J. E. 'Structure and Function in the Living Cell'. *New Biology*, **5**, 26 (1948).

CELL STRUCTURE AND SUB-UNITS

A typical animal cell has, according to Claude, three main sub-units within its surrounding membrane: the nucleus, many rod-like mitochondria, and numerous smaller granules (microsomes) too small to be resolved by the ordinary microscope (Fig. 1). The mitochondria and microsomes are supported in an aqueous phase containing structural proteins, these together constituting the cytoplasm. Amongst the various other particulate inclusions usually present are fat droplets, glycogen, and the much-debated reticular structure known as the Golgi apparatus, while cells with specialized functions like muscle cells or pigmented cells contain muscle fibrils and melanin granules respectively in addition to the other constituents. In plant cells there are often found chloroplasts and other pigment-containing plastids, as well as large vacuoles which contain a solution less viscous but usually more acid than the rest of the cytoplasm. In cells such as those of the liver the

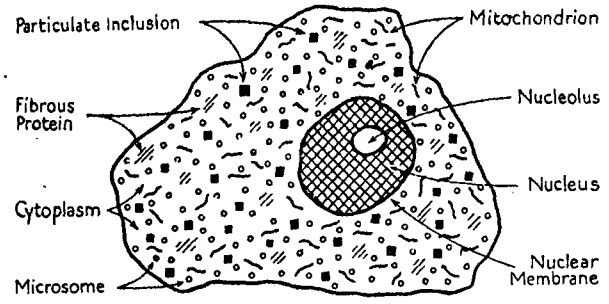


Fig. 1. Diagram of animal cell (schematic).

mitochondria and microsomes may each comprise as much as 15–25 per cent of the total cell weight.

When the cells are disrupted the contents are liberated, and, if suspended in a suitable medium, the separate entities can be examined under the microscope. Such a preparation is known as a homogenate. From a mixture of this kind the particles will gradually settle to the bottom at a rate depending on their size and density compared with that of the liquid. The speed at which the particles settle can be increased by spinning in a centrifuge. When such a homogenate is spun gently the largest and heaviest particles, i.e. the nuclei and unbroken cells that may have survived, come down first, followed by the mitochondria and then the microsomes, leaving the protein and other soluble constituents of the cytoplasm in the supernatant. Higher speeds are necessary to sediment completely the mitochondria and microsomes. The different fractions can be resuspended in a suitable fluid and resedimented until complete separation has been effected. This is the principle of the isolation of cellular components by differential centrifugation.

Homogenization of a tissue may be effected by high speed electric mixers, in which it is minced by a small propeller blade rotating at 10,000 to 15,000 r.p.m., or alternatively by an homogenizer consisting essentially of a glass plunger driven up and down in a tube of slightly greater diameter. The composition of the mixture and its acidity or alkalinity (pH value) are important and must be carefully controlled to prevent loss of labile compounds and to minimize the agglutination of certain particles which readily occurs at certain pH values. The medium selected will, therefore, be determined by the fraction which it is desired to isolate or the special purpose for which the fraction is intended. Some of the most satisfactory solutions at present in use consist essentially of physiological salt solutions or strong solutions of sucrose (cane sugar). When cells are disrupted, changes due to the release of enzymes tend to occur, so that this work must be done at a low temperature in a cold room and the separation processes checked by microscopic examinations at all stages. In this way we can obtain a number of cell fractions with not more than 5 per cent of contaminating impurity.

The technique has been applied in many fields of research, as may be illustrated by a few examples.

(1) It has, for instance, provided evidence that the enzyme systems concerned with the generation of energy in the cell, that is, those concerned with the oxidation of pyruvic acid to carbon dioxide and water, are located in the mitochondria, whereas the enzymes which convert glycogen to pyruvic acid are freely soluble and seem to exist in the soluble fraction of the cytoplasm. This demonstration of different enzymes associated with different cell fractions supports the belief that particular chemical processes are carried out in special parts of the cell.

(2) Studies using isolated fibrils from skeletal muscle are playing an important part in investigating the mechanism of contractile processes. These fibrils, which are the main cytoplasmic components of skeletal muscle cells, show a characteristic striated structure under the electron microscope, and isolated preparations suspended in a medium containing appropriate metal ions will contract on the addition of the coenzyme adenosine triphosphate.

(3) Investigations are being made of the processes involved in photosynthesis using chloroplasts isolated from green leaves. Under suitable conditions such (isolated) chloroplasts can liberate free oxygen on illumination.

(4) Fundamental studies are proceeding in several American laboratories on the production of cancer of the liver in animals. If certain azo-dyes are fed to rats, changes occur which may include those associated with cancer. An examination of the four cellular fractions is being made for the distribution of nucleic acid, protein, etc. It has been claimed that various changes in these fractions have been observed which could not have been detected if only the whole tissue had been analysed.

THE CHEMICAL COMPOSITION OF THE NUCLEUS

While all these sub-units are essential for normal cell metabolism, particularly obvious and serious alterations occur in the cell if the nucleus is damaged, and most cells quickly degenerate and die if the nucleus is destroyed. The nucleus contains the chromosomes which are believed to carry the genes and which most geneticists regard as containing the mechanism for transmitting hereditary characteristics. It has also been credited with being the ultimate source or the control-mechanism of the production of the cytoplasmic enzymes which are responsible for chemical transformations associated with life.

The nucleus has, therefore, been subjected to scrutiny by a wide range of techniques. The methods described above are of particular value, as nuclei can be obtained in quantity with relative ease. Because of these considerations the isolation and the chemistry of cell nuclei have been selected for more detailed discussion.

M. Behrens first obtained cell nuclei some years ago by drying and grinding tissue in the frozen state and then allowing the powdered material to sediment through mixed organic solvents of graded density until the nuclei were separated. Because this method is very laborious, comparatively little work has been attempted with it, though it has the advantage that water-soluble compounds remain in the nuclei. An easier, more general method, introduced by C. A. Stoneburg and developed by A. L. Dounce and others, consists of mincing the tissue in dilute citric acid. This separates the cells and releases the nuclei which can be collected by centrifugation.

More recently, liver cell nuclei have been obtained in a medium containing about 30 per cent sucrose. This gives nuclei which differ in their appearance from those obtained by the other methods and from the nuclei usually observed in stained sections. The latter show typical patterns of localized chromatin. Evidence obtained particularly with the phase-contrast microscope suggests that nuclei in living cells usually have a

homogeneous appearance except for the nucleolus, and that their character in prepared sections, so familiar to the histologist, is due to precipitation during fixation. Such changes can be brought about in nuclei previously isolated with sucrose, by the addition of various salt solutions (Inset 20). This precipitated material has a special affinity for certain basic dyes and consists largely of a protein complex with nucleic acid known as nucleoprotein.*

Isolation of Nucleic Acids

The nucleoproteins were discovered by F. Miescher in 1862 in his classical pioneer work on the nuclei of spermatozoa and pus cells, and he demonstrated that they were composed of nucleic acids and basic proteins. Two classes of nucleic acid are now known to exist in nature; they are usually called pentose nucleic acid and deoxypentose nucleic acid, also known as ribose nucleic acid and deoxyribose nucleic acid.† The former occurs for the most part in the cytoplasm, while the latter is entirely confined to the nucleus. The chemical differences between them will be described later in this article.

The nucleic acids are actually long chain polymers, but those obtained by the earlier workers were usually degraded into smaller molecules owing to the strong alkali used in their extraction, and it is only within the last decade or so that satisfactory macro-molecular preparations have become readily accessible, largely as a result of observations by A. E. Mirsky and A. W. Pollister in New York, first published in 1942. This work also led to much better understanding of the location of these components. They found that when minced animal tissues were extracted with physiological saline (0.85% NaCl) most of the soluble material outside the nucleus was washed out of the cells. This included pentose nucleoproteins which could be

* As explained later, the *nucleoproteins* are proteins, e.g. histones, combined with nucleic acid and occur in both the nucleus and the cytoplasm of cells. In contrast, any proteins present in the nucleus are nuclear proteins.

† The reason for these alternative descriptions will appear later.

precipitated from the extract with dilute acid and redissolved in dilute alkali. Extraction of the residue, which is essentially composed of nuclear material, with stronger saline (5% or more) gave a solution with a very high viscosity. When this was diluted with 6 volumes of water so that the final concentration of salt was physiological (0.85%) a white fibrous precipitate of deoxypentose nucleoprotein was formed. This material can be redissolved in strong salt solution and reprecipitated any number of times, and in this way it can be freed from contaminants.

Nucleic acids can be obtained from such nucleoproteins by several methods. Thus the class called deoxypentose nucleic acid (DNA) can be obtained by salting out the protein component by the addition of solid sodium chloride, or alternatively by shaking the nucleoprotein (NP) solution with chloroform containing an alcohol such as butanol. This mixture is immiscible with the NP solution and an emulsion is produced. The protein which is split off from the complex forms a gel with the chloroform-butanol while the nucleic acid remains in the aqueous part of the solution. This gel is removed by centrifuging and the process is repeated until the nucleic acid is judged free from protein. The nucleic acid solution is then poured into alcohol when a fibrous white precipitate of DNA is formed (Inset 21). Pentose nucleic acid (PNA) can be obtained similarly in an undegraded condition by shaking a solution of pentose nucleoprotein with the chloroform-alcohol mixture, by denaturing the protein* by adding alcohol, or by heating and then extracting the PNA with 10 per cent NaCl.

Structure of Nucleic Acids

Investigation of the nucleic acids shows that they are built up of much smaller units – namely the purine and pyrimidine bases, a sugar component, and phosphoric acid molecules. The last two components are familiar, at least in their general description. The purines and pyrimidines are structurally related, the

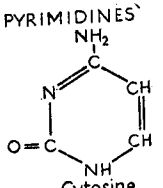
*See following section on *The Proteins of the Nucleus.*

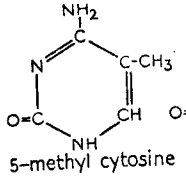
latter having a single ring containing 4 carbon atoms and 2 nitrogen atoms, and the purines having in addition an iminazole ring built onto it, with 2 extra nitrogen and 1 extra carbon atoms.

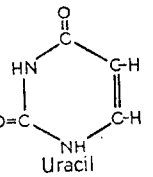
The components found in nucleic acids are:

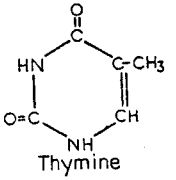
Bases

PYRIMIDINES

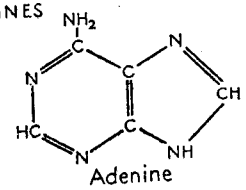

Cytosine

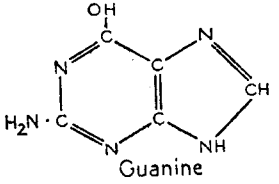

5-methyl cytosine


Uracil


Thymine

PURINES

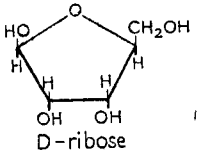

Adenine


Guanine

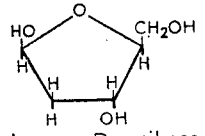
Sugars, believed to be

$C_5H_{10}O_5$

ring form as believed to exist in nucleic acid molecule

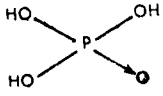

D-ribose

$C_5H_{10}O_4$


2-deoxy-D-ribose

Phosphoric acid

H_3PO_4



The two groups of nucleic acids which occur in nature are usually distinguished by their sugar component. Those in which the sugar is a pentose are termed pentose nucleic acids, and those containing 2-deoxypentose,* deoxypentose nucleic acids. In a number of specimens the sugars have been identified as D-ribose and 2-deoxy-D-ribose respectively, and there is no evidence for the occurrence of any sugars except these two. Another difference between the two classes of nucleic acid lies in the pyrimidine units. In DNA these are thymine and cytosine while in PNA uracil is found along with cytosine. The components found in the nucleic acids are summarized below in Table I.

TABLE I: *Composition of the Nucleic Acids*

Unit	PNA	DNA
Sugar	Pentose (Believed to be D-ribose)	Deoxypentose (Believed to be 2-deoxy-D-ribose)
Bases Purine Pyrimidines	 Adenine Guanine Cytosine Uracil	 Adenine Guanine Cytosine Thymine 5-methyl cytosine (in some cases)
Acid	Phosphoric acid	Phosphoric acid

When a nucleic acid is boiled with strong acid it is hydrolysed into its component molecules (2-deoxy-D-ribose is destroyed) and a mixture of purines and pyrimidines is produced which can be separated by fractional crystallization, formation

*The prefix 2-deoxy denotes the fact that deoxy pentose could be obtained in theory by removing an oxygen atom at position 2 of the ring. Sugars, in general, exist in alternative forms which are mirror images of one another, and their configuration in space is denoted by D or L respectively.

of insoluble compounds, etc. In small quantities the separation can be accomplished most satisfactorily by the recently developed technique of chromatography on filter paper.* When solutions of acidified alcohols such as butanol or isopropanol are allowed to flow over a spot of the solution of purines and pyrimidines dried on the paper (containing about 50 microgramme of the constituents) these are carried forward distances which depend largely on the character and size of the molecules. Various solvents can be selected so that complex mixtures can be separated with great accuracy. In the case of purines and pyrimidines, the positions which the separated substances finally occupy in the form of spots on the paper-strip are detected by placing photographic paper under the chromatogram and exposing it to ultra-violet light of suitable wave-length. The paper will transmit the 'light' whereas the deposited purines and pyrimidines will absorb it, leaving the sensitive photographic paper relatively unaffected where they are present. When the paper is developed, these spots will show white against a black background (Inset 22). When the deposits have been located, the several areas can be cut out of the chromatogram, extracted with dilute acid and the individual bases estimated by measuring with a spectrophotometer the extent to which ultra-violet radiation of particular wave-lengths is absorbed by the extracts. This enables a very accurate analysis to be carried out, for example on a few milligrammes of dried nucleic acid.

The compositions of some typical nucleic acids are given in Table II (*overleaf*). One of the recent accomplishments of chromatography has been to demonstrate that the proportion of the four main bases in the molecule varies widely with the source of the nucleic acids. Another pyrimidine constituent, 5-methyl cytosine, has also been shown to be present in relatively small amounts in nucleic acids obtained from a number of plant and animal tissues.

* See Lester Smith, E. 'Chromatography'. *Science News*, **13**, 9 (1949).

TABLE II: *Proportions of Purines and Pyrimidines in Nucleic Acids from Various Sources*

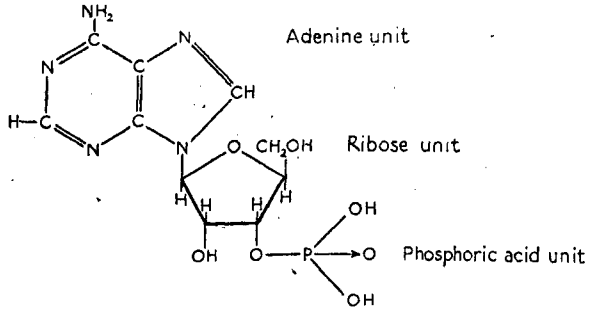
(Adenine = 1·00 in all specimens)

Source	Guanine	Cytosine	Uracil	Thymine	5-methyl cytosine
Calf thymus DNA	0·77	0·76	0·0	1·0	0·05
Calf liver DNA	0·89	0·85	0·0	1·04	—
Ox spleen DNA	0·77	0·76	0·0	1·0	—
Calf liver PNA from cytoplasm	1·83	1·22	1·12	0·0	—
Calf thymus PNA from cytoplasm	1·88	1·44	0·89	0·0	—
Calf liver PNA	1·0	0·39	0·45	0·0	—
Rat bone marrow cell DNA	0·77	0·74	0·0	1·0	0·04
Wheat germ DNA	0·88	0·66	0·0	1·0	0·22
Mouse tumour cell DNA	0·94	0·72	0·0	0·88	—
Yeast PNA	1·13	0·71	0·73	0·0	—
Sea Urchin sperm DNA	0·70	0·64	0·0	0·91	—
Mycobacterium tuberculosis PNA	1·86	1·92	0·31	0·56	—
Bushy stunt tomato virus PNA	1·125	0·88	1·01	0·0	—

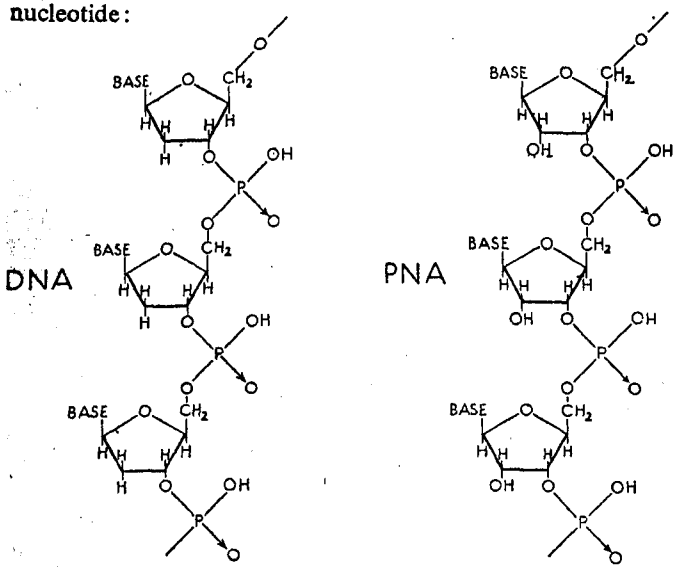
All figures are proportional to 1 Gram Molecule of Adenine

The large nucleic acid molecule is broken down by dilute mineral acids or by certain enzymes into smaller units, the nucleotides, which consist of one of the bases linked to a sugar residue and a phosphoric acid molecule. A typical nucleotide

from PNA is adenylic acid, and the linkages of the units are shown in the diagram below:



The complete nucleic acid molecules are generally considered as being built up of nucleotides linked through the hydroxyl (OH) groups of the sugars and the acid groups of the phosphoric acid with elimination of water molecules, each unit of the latter acting as a bridge joining the sugar units of each nucleotide:



Evidence from physical measurements lead us to believe that as many as a thousand such nucleotides may be contained in a macro-molecule of nucleic acid. The links depicted in the above formulae seem to form the majority of those present, though evidence is accumulating that other bonds such as those involving branched chains are present, and these must play a very important part in determining the structure of these macro-molecules. No information is yet available about the sequence of the bases in nucleic acids nor is there any indication whether a nucleic acid from a particular source consists of one type of molecule or of many closely related structures.

In solution the nucleic acids form a long thread-like molecule to which the high viscosity of the solution is due. DNA can be obtained with a molecular weight of the order of 1–3 millions, but an exact value is very difficult to estimate, for it may vary from preparation to preparation and its accurate evaluation is complicated by the tendency of the molecules to aggregate. PNA has a somewhat smaller molecule with molecular weights of about 10,000–300,000.

In the cell cytoplasm PNA seems to exist mainly in the microsomes and mitochondria in association with proteins and fat-like substances. As indicated above, DNA is found in all cell nuclei of animals and plants, and in the equivalent regions of bacteria and other lower organisms. Certain viruses of animal origin also contain DNA. PNA is distributed in both the nuclei and cytoplasm of cells, and forms part of many types of plant viruses. Some of the plant viruses have, in fact, been obtained as crystalline pentose nucleoproteins.

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 PNA 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 produce large amounts of protein-containing secretion. Gastric mucosa cells which make

the protein enzyme, pepsin, contain much PNA whilst neighbouring cells which secrete only hydrochloric acid have very little.

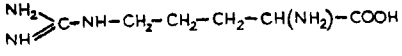
THE PROTEINS OF THE NUCLEUS

Our knowledge about the protein constituents* of the cell nucleus is much less detailed than that about the nucleic acids. This is largely due to the difficulties attending both the isolation and the analysis of such structures. Fuller understanding of their localization and their intimate structure is, however, particularly necessary since the biological specificity of the proteins is mainly determined by the manner and proportion in which their constituent amino-acids are arranged.

The nucleus is the only site where the very simplest kinds of protein, the protamines and the histones, are found in quantity.

(a) Basic Proteins

A basic, nitrogen-containing substance was isolated from pus cells by Friedrich Miescher in Germany about 1868, but its relation to the proteins was not fully elucidated until nearly twenty years later when he obtained from salmon sperm a basic substance which he named 'protamine'. We now know that the protamines are the simplest of a series of compounds, the most complex of which are the typical proteins such as occur in blood serum, egg white, or muscle. The protamines are probably constructed in the form of a single chain of amino-acids – called a polypeptide chain – and about two-thirds of the molecule consists of the amino-acid arginine:



to which its basic nature is chiefly attributable. The complete molecule contains some 4 to 7 other amino-acids, e.g. histidine,

* See Kendrew, J. G. 'The Structure of Proteins'. *Science News*, **11**, 125 (1949).

lysine, valine, proline, and tyrosine, has a molecular weight of 2,000 to 8,000, and corresponds with a structure containing 20 to 80 amino-acid residues (Table III). It can combine with nucleic acid to give a nucleoprotein, in which combination it probably exists during certain phases of cell division.

Such a nucleoprotein was isolated by Miescher when working on the spermatozoa of the Rhine salmon in his laboratory at Basel, and from it he separated a compound containing an organic base (protamine) and the organic acid in which phosphorus was present which is now known as nucleic acid. He observed that during their ascent of the Rhine from the sea to the spawning beds, the muscle tissues of the salmon greatly decreased, while the reproductive organs, producing eggs or spermatozoa, enormously increased although little or no food was consumed during this period. The head of the spermatozoa may be regarded as a metamorphosed nucleus, so it was concluded that the very high content of protamine had been produced from material derived from muscle tissue. Subsequently the sperm cells of certain other species of fish were found by A. Kossel to be rich in basic polypeptides. Following his suggestion, the term protamine has been assigned to this whole group of simple proteins, and the various individual types which are distinguished are named after the source from which they are obtained, e.g. clupeine from herring sperm, scombrine from the mackerel, salmine from salmon.

It is of historic interest to note that, by selecting sperm heads, Miescher anticipated in a simple way the more recently developed technique using isolated nuclei from cells of tissues. Also in his early work on pus cells, he removed the cytoplasm by extracting them with a dilute solution of sodium sulphate and digesting them with an artificial gastric juice. He thus used essentially the same kind of material that we use to-day, starting with tissue homogenates.

In 1884 Kossel described a basic protein material derived from the red blood corpuscles of geese which, in common with other bird erythrocytes, possess nuclei. This he named 'histone'. He saw certain similarities in its properties to

Miescher's base from salmon sperm and proposed the name protamine for that group of substances. Histones contain a greater number and variety of amino-acids than do protamines and are more complex therefore, but they are probably built up in a very similar way. They do not contain as high a percentage of the more basic amino-acids, however, and are therefore more nearly related to the typical proteins. Their estimated molecular weight is about 15,000. Histones are more widely distributed than the protamines and are present in appreciable amounts in the nuclei of the tissue cells of all animal species examined. From the unripe spermatozoa of certain fish, and from the ripe spermatozoa of others such as the cod fish, substances have been prepared which also can be classed as histones. Mammalian sperm, on the other hand, do not contain histones, and the proteins (and also the nucleic acid) which they contain possess solubility characteristics which differ from those present in the nuclei of the mammalian body tissues.

It has been reported that transitional forms between protamines and histones are present in the sperm nuclei of some invertebrates, and one such substance has been obtained very recently from fowl sperm.

The particular histone which Kossel obtained from bird blood-corpuscles is extracted by dilute hydrochloric acid, is soluble in water, and is precipitated by ammonia. The variety present in the cells of such organs as the thymus can be obtained by extracting the nuclei with dilute sulphuric acid and then precipitating it from the solution by very cautious addition of dilute alkali and drying the material so obtained with alcohol, or better, by evaporation from the frozen state at a very low temperature in a high vacuum. This minimizes the structural changes (known as denaturation) which occur in many kinds of proteins when they are exposed to heat or alcohol. The product obtained by this latter method remains easily soluble in water, whereas denatured proteins tend to become insoluble.

In order to determine whether any particular protein may

properly be called histone we rely principally on examining the general proportions of its constituent amino-acids (Table III), the size of its molecule and its solubility characteristics. The class of histones in general is, as described above, *insoluble* in dilute alkalis.

TABLE III: *Composition of Basic Proteins*

Amino-acid	Amino-acid nitrogen values as percentage of total protein nitrogen content	
	Protamine Salmine (from salmon sperm)	Histone (from calf thymus)
Alanine	0·6	6·0
Ammonia	0·0	4·8
Arginine	89·5	30·7
Aspartic acid	0·0	3·3
Glutamic acid	0·0	2·2
Glycine	1·8	5·2
Histidine	0·0	4·0
Isoleucine	0·0	12·0
Leucine	0·6	3·0
Lysine	0·0	10·8
Phenylalanine	0·0	1·9
Proline	2·3	2·7
Serine	3·9	3·4
Threonine	0·0	3·1
Tyrosine	0·0	1·4
Valine	1·2	4·9
Total	99·9%	99·4%

Basic proteins are probably joined by a salt link between the phosphoric acid groups of the nucleic acid and the guanidine and amino groups of the basic amino acids, which is analogous to the formation of ammonium salts of simple acids. If neutral solutions of DNA and histone are mixed, an insoluble nucleoprotein complex is formed in which the large multicharged ions are held together by purely electrostatic attractions. The size of the DNA molecule is such that in the

nucleoprotein complex several molecules of histone or protamine are combined with one of DNA.

(b) Other types of Nuclear Proteins

In addition to the alkali-insoluble histones, it has been shown that the body cell nuclei contain large quantities of a protein which is *soluble* in dilute alkali. Also the presence of two other less well defined protein fractions has been reported but practically nothing is yet known about their structure.

It is surprising in view of the development of the gene theory of heredity that components other than nucleic acids and simple basic proteins were not recognized in cell nuclei until D. T. Mayer in Missouri found other protein fractions in calf thymus nuclei, first reported in 1939. In 1943, Stedman and Stedman working in Edinburgh reported the presence in a variety of nuclei of considerable amounts of a non-basic protein which they called 'chromosomin'. This may be related to an alkali-soluble fraction obtained by various investigators. Some work is now proceeding on these proteins by a number of investigators in this country, in America, and in Russia, but comparatively little is yet known about their constitution.

Less still is known about the composition of the protein components of PNA-protein complexes. Those which have so far been studied, derived from certain plant viruses, show no exceptional features.

QUANTITATIVE DATA ON CHEMICAL COMPONENTS OF NUCLEI

The difficult task of carrying out complete determinations of the quantity of nucleic acid and the various types of protein components in the nuclei from various species and organs has been attempted by only a few biochemists and it is not easy to compare the results obtained under various experimental conditions. Some of the values reported, however, have been set out in Table IV. Attention has been called by several workers to the relatively large amount of PNA in the liver nuclei as

TABLE IV: *Chemical Components of Nuclei* (Percentages)

Source of Nuclei	DNA	PNA	Histone	Other Protein	Lipid
Calf thymus	38	*c.* 0·5	29	33	—
Calf thymus 'chromosomes'	37	1·0	55	8	—
Mouse liver	27	3·4	66*		3·4
Fowl erythrocytes	43	—	24	33	—
Rat liver	13·9	6·2	—	—	—
Mouse spleen	28	1·3	70*		—
Human thymus	37	—	27	*c.* 25	—
Human liver	27	—	20	*c.* 40	—
Herring sperm head	59	—	22†	*c.* 10	—

* Total protein † Protamine *c.* – about

compared with thymus or spleen. This is illustrated in the above table, as is the fact that the chromosomes themselves are composed mainly of nucleic acid and basic protein with only a small amount of other protein. This latter might be regarded as being chiefly in the less organized part of the nucleus or nuclear sap in which the chromosomes are embedded.

THE CHEMISTRY OF THE CHROMOSOMES

Some of the most important and interesting problems in cell chemistry are connected with the composition of the chromosomes themselves.* The nucleus is generally looked upon as consisting of chromosomes embedded in a nuclear sap and enclosed by a nuclear membrane. Most workers agree that the backbone of the chromosomes consists of a protein 'fibre' and

* See Callan, H. G. 'Chromosomes'. *New Biology*, 7, 70 (1949).

the DNA is usually regarded as being an essential part of the chromosome structure. The Stedmans, however, maintain that the chromosomes consist of an acidic protein 'chromosomin' which is embedded in a gel of nucleohistone.

Long before DNA was extracted from isolated nuclei it was demonstrated in this locality by a special staining technique known as the Feulgen reaction considered to be specific for DNA. In this test, the tissue is submitted to the action of dilute hydrochloric acid at a temperature of 60° C. and is then treated with a solution of fuchsin which has been first rendered colourless by reduction with sulphur dioxide (leuco-fuchsin). Any structure containing DNA becomes stained a purple colour. Because some criticism had been directed at this important cytochemical test, a detailed examination of the Feulgen reaction has been made, particularly by M. Stacey and his collaborators at Birmingham; it must be considered that, when correctly applied, this test does, in fact, demonstrate the location of DNA in chromosomes.

Nucleic acids are characterized by an intense absorption of ultra-violet light particularly at a wave-length of 2700 A, a property which can be used to map out the position occupied by the nucleic acids in suitably prepared sections of tissue. It is not possible to differentiate between DNA and PNA by this method, since the absorption is due to the purine and pyrimidine units which occur in both. The ultra-violet microscope is a very sensitive instrument which has been developed by T. Caspersson and his co-workers in Stockholm. They have utilized it to make quantitative analyses of the nucleic acids and proteins in various cells by measuring their ultra-violet absorption *in situ*. Some of the quantitative results obtained by this method, particularly with regard to the distribution of protein in cells, may need some revision because the ultra-violet absorption of proteins (due mainly to tyrosine and tryptophane) is only about 1/80th of that of the same amount of nucleic acid, so that the absorption by a nucleic acid is likely to obscure that of the protein. Also, little is known of the amino-acid composition of cell proteins and their absorption,

attributed to the aromatic rings, has a maximum at 2,900A, which is very near that of the nucleic acids. Precise calculations of the protein content by this method, therefore, present many difficulties, but the above techniques have proved of great value for the localization of DNA in metaphase chromosomes. The large salivary glands of certain insects (e.g. fruit fly) have been extensively used for studies involving these techniques and provide specially suitable material because of the giant chromosomes which they contain. The nuclei persist in a sort of permanent prophase, and the chromosomes have bands of DNA at definite points along their structures. The distribution of DNA in the usual type of inter-phase nucleus is not known.*

Isolated Chromosomes

If cell nuclei are subjected to violent mechanical disruption in a high-speed mixer, they in turn can be fragmented, and microscopic threads can be collected in quantity and purified by differential centrifugation. Mirsky and his associates claim that these threads consist of liberated chromosomes because of their microscopic appearance and staining properties. This belief has been criticized by some investigators and supported by others who have examined the threads by means of the electron microscope and by micromanipulation. The gross chemical composition of these 'separated chromosomes' is very similar to that of separated nuclei, for they contain DNA, PNA, histone, and an alkali-soluble protein (Table III).

CHEMISTRY AND GENETICS

It is possible that DNA may enter into the genetic control of the cells of the higher organisms, but at the present time speculation concerning the chemical structure of the gene would be hazardous in view of our scanty knowledge of the intimate structure of nucleoproteins and similar lack of knowledge

*For details of the characteristic stages of cell division, see Callan, H. G. 'Chromosomes'. *New Biology*, 7, 70 (1949).

regarding other chemical constituents. A remarkable property of purified DNA extracted from certain bacteria, has however, been shown by O. T. Avery, C. M. MacLeod, and M. McCarthy. They prepared DNA in a pure form from Type III pneumococci which could effect the transformation of uncapsulated 'rough' cells of pneumococci Type II into capsulated 'smooth' cells of pneumococci Type III. Thus a specific hereditary change was induced in this organism which might be believed to be an example of a chemically directed mutation. Since the original observation of Avery and his colleagues similar 'mutations' have been described for other bacteria.

SOME GENERAL CONCLUSIONS

By using such recent techniques as are outlined above, a good many data on the content and composition of the nucleic acid in cells from various tissues of many organisms have been collected. These results enable us to draw a number of general conclusions.

The DNA isolated from different organs of the same animal species seems to be constant in composition with respect to the relative amounts of purine and pyrimidine bases but differs from samples obtained from other species. On the other hand, PNA derived from different organs of the same animal appears to have a different composition and to vary even between the nucleus and cytoplasm of the same cell. It is possible that some of the proteins in cell nuclei from different organs of the same animal also have small but distinct qualitative differences in composition.

A second important conclusion was drawn in 1948 by A. Boivin, R. Vendrely, and C. Vendrely, of Strasbourg, who produced evidence that a given set of chromosomes is associated with a constant amount of DNA. They prepared a homogeneous suspension of isolated nuclei and divided it into two portions. In one of these they estimated the total quantity of DNA, and in the other counted the number of nuclei under the microscope in the same way that blood cells are enumerated. Thus it was possible to estimate the average DNA content

per nucleus. It was found that nuclei containing two sets of chromosomes (diploid) from the various organs of an animal of any particular species had about the same DNA content, e.g. 6×10^{-12} g. of DNA per nucleus for beef tissues. This was exactly twice the DNA content of the bull spermatozoa nucleus which contains only one set of chromosomes (haploid). The DNA content of diploid nuclei was observed to vary widely from species to species.

When polyploid nuclei, i.e. those containing multiple sets of chromosomes, were subsequently examined, the DNA content appeared to be a simple multiple of the value for the haploid nucleus. Thus in a tissue like liver there are present nuclei with two (diploid), four (tetraploid), and eight (octaploid) sets of chromosomes, and the DNA content as determined chemically gives an average value. Estimates of the relative DNA contents of these different types of liver nuclei have also been made using cytological methods and these support the idea of a constant value of the DNA per haploid set of chromosomes. It seems necessary to conclude that the additional nucleic acid required for the duplication of the chromosomes is synthesized immediately before division.

Studies employing radioactive isotopes of carbon and nitrogen show that the body proteins are continually changing their component amino-acids, i.e. they exist in a state of dynamic equilibrium. Similarly some of the PNA molecules seem to undergo active exchange reactions with simpler molecules. The turnover of the nuclear DNA is, however, much less, and this molecule seems relatively stable, at least with respect to its turnover of phosphorus atoms.

The above observations, viz. constancy in amount, constancy in composition in any particular type of cell, and metabolic inertness, suggest that DNA has an important stabilizing role in the chemical basis of heredity. This is in striking contrast to the variations in quantity and composition of the PNA.

Such facts as the constancy of the DNA, variability of the PNA (nuclear and cytoplasmic) and the variability of the nuclear proteins are the first fruits of researches into the

chemistry of cells which in due course may enable us to understand more of the mechanism of the differentiation of the single-celled fertilized egg into the many types of cells, with such diversity of function, which comprise the fully developed animal. An attempt has already been made to compare the DNA content of the nuclei of a wide range of organisms and to speculate what this may mean from the standpoint of their relationship in evolution. Many biochemists anticipate that studies of this kind will throw light on the differences between the behaviour of the cells found in normal tissues and those present in malignant tumours.

In this survey of a field of biochemistry which is in active development many aspects have been omitted, but it may suffice to indicate that chemical investigations with this approach have wide implications in many branches of biology. M. J. D. White (1951) has expressed the opinion that 'Classical genetics and cytology have gone about as far as they are capable in explaining the gene: it remains for the chemists to complete the analysis of . . . the irreducible units of living structures.'

FURTHER READING

Davidson, J. N. *The Biochemistry of the Nucleic Acids.* London, Methuen (1950).

Schneider, W. C., and Hogeboom, G. H. 'Cytochemical Studies of Mammalian Tissues: the Isolation of Cell Components by Differential Centrifugation'. *Cancer Research,* **11,** 1 (1951).

Caspersson, T. *Cell Structure and Cell Function.* Norton & Co. Inc. N.Y. (1950).

Claude, A. 'Proteins, Lipids, and Nucleic Acids in Cell Structures and Functions'. *Advances in Protein Chemistry,* **5,** 423 (1949).

Cold Spring Harbor Symposium on Quantitative Biology, 12: Nucleic Acids and Nucleoproteins. Long Island Biological Association, Inc., 1947.

Symposia of the Society for Experimental Biology, No. 1: Nucleic Acid. Cambridge University Press, 1947.

DOUBLE STARS

W. H. MARSHALL

IT is about 170 years since Sir William Herschel started the systematic study of double stars. Since then, the study of these objects has shown steady progress, punctuated by intervals of apparent stagnation when there seemed to be no more to be done, and by periods of 'rebirth' under the influence of new instruments or new ideas. We need look back only a few years to find a stagnant period, when the future of double-star astronomy seemed to some to offer no prospect beyond the careful checking of old observations. And we are still at the beginning of a new era, when double and multiple stars seem destined to play an increasingly important, perhaps a decisive, role in the formulation of our ideas on the evolution of the stars, and of the universe itself. The time seems, therefore, appropriate for a review of our knowledge of binary stars, how it has been obtained, what information may be derived from it, and where it seems to be leading.

A double star, or binary, is a system consisting of two stars revolving around a common centre of gravity under the influence of their mutual gravitational attraction. Since the distance of the stars from us is much greater than the distance at which gravitation would cease to hold the stars together, it is clear that the members of a binary system must appear near to each other in the sky. But two stars which happen to lie almost in the same direction will appear thus close together, although they may actually be very far apart in space. The first task in the study of double stars is therefore to eliminate these spurious systems or 'optical doubles' as they are called. This can be done by measuring the proper motions.

The most easily observed motion of the stars, from east to west across the sky, is, of course, only an *apparent* motion, a

reflexion of the rotation of the earth from which we observe it; and it produces no change in the positions of the stars, relative to one another. Accurate measurements show, however, that the stars do have relative motions of their own, their so-called proper motions, and these can be measured against the background of stars too far away for their movements to be observable. It is extremely improbable that two unconnected stars will have the same proper motion, and so the components of an optical double will be found to be moving independently. They are thus distinguished from true binaries, the proper motions of whose components differ only by a small amount due to the orbital motion.

The spurious systems being thus eliminated, we are left with the true binaries, which may be divided into three overlapping groups: visual binaries, spectroscopic binaries, and eclipsing binaries. The names arise, as will be seen, from the methods used to study them.

VISUAL BINARIES

Visual binaries are those systems in which the stars are far enough apart in the sky to be separated by a telescope. They were the first type to be discovered and they are studied with the filar micrometer, whose principle is shown in Fig. 1. The brighter star of the pair is brought to the centre of the field of view of the telescope, as shown, and the drive of the telescope

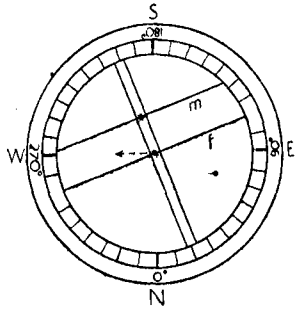


Fig. 1. Principle of filar micrometer (see text).

is stopped. This causes the image of the star to drift across the field from east to west as the earth rotates; the speed of this drift is one of the most surprising things noticed by a beginner at the telescope. The outer circle of the micrometer is then turned to the position shown, with 270° marking the west point.

With the circle set, the double wire is now turned to enclose the two stars as shown, and the *position-angle* (*p*) of the pair is read from the scale. This is defined as the angle between the north direction and a line drawn from the brighter component to the fainter; and it is measured from north through east, right round the circle. In Fig. 1 it is about 200°. When the principal star is central in the field of view, it lies on the fixed wire, *f*, of the other pair. The movable wire, *m*, is now set over the fainter star by means of a micrometer screw, whose head is graduated to show the corresponding distance, *d*, in seconds of arc.

These measurements are repeated at intervals depending on the apparent speed with which the binary changes, and when a sufficient number of observations has accumulated, the apparent orbit can be drawn. A point is taken to represent the principal component, and through it are drawn axes, as shown in Fig. 2, giving position angles 0°, 90°, 180°, and 270°. Each

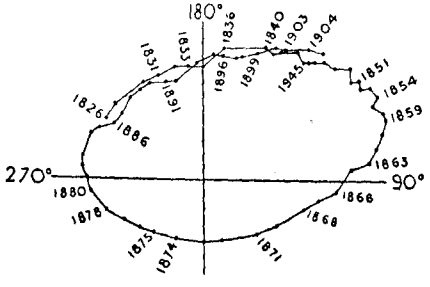


Fig. 2. Orbit of ζ Ursae Majoris (from Burnham's *General Catalogue of Double Stars*, by courtesy of the Carnegie Institution of Washington.

observation is then represented by a line which makes the appropriate angle with the axes, and whose length represents the distance of the components on a convenient scale. Fig. 2 shows a set of actual observations plotted in this way, and it is clear that, when smoothed out to allow for errors of observation, the apparent orbit of the fainter star around the brighter one is an ellipse.

The apparent orbit is not, however, the true orbit, unless its plane happens to be perpendicular to the line of sight, which it very rarely is. The true orbit differs in eccentricity from our oblique view of it, and the principal star lies at one focus. The problem of deriving the true orbit from the apparent one is a matter of geometry, and need not concern us here.

The most important elements of the orbit are: *P* the period of revolution; *e* the eccentricity, which measures the elongation of the ellipse; *a* the semi-major axis, i.e. half the length of the ellipse; and *i* the inclination of the plane of the orbit to a reference plane, perpendicular to the line of sight, and called the plane of the sky.

For fairly wide pairs, the measurements may be made photographically. In this case, the pair is photographed at intervals, and the position-angle and distance are measured from the photographic plates and plotted as before. Two photographs of the binary called Krüger 60 are shown in Inset 23, where the change in the system can be easily seen. The method cannot be applied to the closest doubles. Because of the nature of the photographic process, the images of the stars are always spread over a finite area; and, if the separation is less than about 2″, the images overlap and become difficult or impossible to separate. On the other hand, the photographic method has two advantages, compared with the micrometer method. First, it is less tedious, since the most trying part of the work, the accurate measurement of angles and distances, is done in the comparative comfort of the laboratory. And second, it is often possible to measure the displacement of *both* components of the binary relative to the background of distant stars. This enables the orbits of both stars about the centre of gravity to be determined,

instead of only the relative orbit of the fainter about the brighter. As we shall see later, this information is of great value.

It is not only the photographic method of measuring visual binaries that is limited in its application. The micrometer method also is confined to those pairs whose components can be separated by the telescope, whose resolving power is limited to a value given by Dawes' formula $R=4.56/D$ seconds of arc, where D is the diameter in inches of the object glass of the telescope. The values of R given in Table I are the *smallest* separations which we can hope to measure with telescopes of the stated sizes; and this can be achieved only when the two stars do not differ much in brightness. Otherwise, the fainter component becomes very difficult to see in the glare of the brighter.

TABLE I: *Size and Resolving Power of Telescopes*

D. ins	10	20	30	40	60	80	100
R. secs	.46	.23	.15	.11	.076	.057	.046

The Interferometer

The only way to extend measurements beyond the limit shown is by increasing the resolving power of the apparatus. This can be done by the use of an interferometer. If light from a single source S (Fig. 3*a*) is split into two beams by the slits A and B, and allowed to recombine at C, we observe a series of bright and dark bands, formed by interference of the two beams (Fig. 3 *b*). These interference fringes are separated by a distance $x=f\lambda/2D$, where λ is the wavelength of the light used; and the central fringe is always bright.

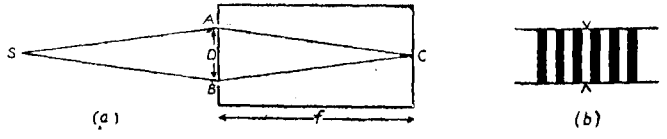


Fig. 3. Principle of the interferometer (see text).

Now light from one source cannot be made to interfere with light from another, so when two sources are used, the two systems of fringes are formed quite independently of each other. If, then, the slits are placed over the objective of a telescope of focal length f, we see in the focal plane a double set of fringes overlapping each other to some extent. The distance y between the central fringes of these two systems is the distance between the centres of the images formed by the telescope without the slits, and we have $y=f\alpha$, where α is the (small) angle between the two stars. If, then, we set the slits so that the line joining them is parallel to that joining the two stars, and adjust D so that $y=x$, the bright fringes of one star will fall on the dark fringes of the other, and the fringes will disappear. Their disappearance thus shows that $y=x$, or, from the formulae given, that $\alpha=\lambda/2d$ where d is the critical separation.

The resolving power of the interferometer is about twice that of a telescope of diameter equal to the distance between the slits. By a variation of the construction, in which mirrors were carried on a 20-ft. beam attached to the 100-in. reflector at Mount Wilson, resolution equal to that of a 400-in. telescope can be obtained. In this way, double stars with separations of only 0″.011 have been measured.

The great advantage of the interferometer is that it bridges the gap between the visual binaries and those which can be detected and measured only by other methods. There are, however, many thousands of pairs still beyond its reach, and the first type we have to mention is the spectroscopic type.

SPECTROSCOPIC BINARIES

The lines in a star's spectrum may be altered in position, as compared with those in a terrestrial spectrum, by motion of the star in the line of sight. And they may become double or multiple, when seen with high resolution, under the influence of a magnetic field. But in 1889 E. C. Pickering, at Harvard, discovered that the lines in the spectrum of Mizar (the brighter component of the double star ζ Ursae Majoris in the tail of

the Great Bear) were sometimes double and sometimes single. A typical set of spectra is shown in Inset 24.

The changes observed are explained by the fact that the apparently single star really consists of two components revolving around each other, although they cannot be seen separately. If, in Fig. 4, A and B are the two stars, B being assumed to move around A as shown, we see that at B_1 and B_3, when B is moving across the line of sight, the spectral lines of the two stars coincide, and so we see a single line. At intermediate positions, however, the velocity of B has a component along the line of sight, and so its lines will be shifted by the Doppler effect, and will be separated from those of A, which remain stationary. The displacement will be a maximum towards the red at B_2 and a maximum towards the violet at B_4. In actual fact, of course, A will also be moving around the common centre of gravity, so that its lines will also move back and forth through their mean positions, and the result will be an increase in the observed separation of the lines.

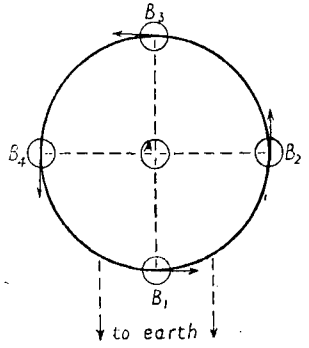


Fig. 4. Observation of spectroscopic binaries.

From the spectra we derive a curve showing the velocity at different times during the period of revolution; and from the shape of this velocity curve we can derive the elements of the orbit, with the exception of the inclination i.

When only one of the components is bright enough to give a detectable spectrum, what we observe is an oscillation of one set of lines about its mean position. In this case, we cannot derive so much information from the observations. When both spectra are available, we get both orbits, and hence the masses of the components, as we shall later see. From one spectrum, we get only one orbit and cannot estimate the masses. But additional information can be derived from those spectroscopic binaries, such as Mizar, which have also been measured with the interferometer.

ECLIPSING BINARIES

The third type of double star, and in some ways the most important, is the eclipsing type. It is recognized by a regular variation, of a distinctive pattern, in its brightness. As shown schematically in Fig. 5*a*, the light of the star remains constant for some time, then fades rapidly to a minimum, from which it rises again to its steady value. This dimming occurs at regular intervals. Sometimes, as in Fig. 5*b*, there is a second, smaller dip in the light-curve, so that the curve then contains a primary and a secondary minimum, each recurring constantly at the same interval.

Fig. 5. Light curve of eclipsing binary (diagrammatic).

This behaviour is explained by the periodic eclipsing of the stars by each other. If the line from the earth to the star lies nearly in the plane of the orbit, it is clear that once in every revolution each star will come approximately between us and the other. Each of the components therefore undergoes an eclipse, total if the orbit is exactly edge-on to us, otherwise partial. If one of the stars is brighter than the other, its eclipse will reduce the total light more than that of its companion, so

that the form of the light-curve can be completely explained. The period of light-variation is, of course, the period of revolution of the stars in their relative orbit.

The construction of the light-curve of an eclipsing binary depends on measurements of the brightness of the system, which, we must remember, is seen as a single source of light. The measurements may be made by any of the customary methods, visual, photographic or photoelectric. One method, which is particularly useful for short-period systems, is worth describing. The photographic plate is placed slightly beyond the focal plane of the telescope, so that the image formed of any star is a small, uniformly illuminated disc, instead of the usual diffraction disc. The plate is then exposed a number of times, being moved between exposures, so that successive images do not overlap. Fig. 6 shows a typical result. This technique is a good one for finding stars of variable brightness, and the light curve which can be drawn from the observations will indicate quite clearly whether the variation is of the type due to eclipses. For really accurate determinations of the period, and of the range of the variation, however, the best method is the photoelectric one. This method, especially when the new electron-multiplier type of cell is used, gives results which cannot be bettered by any other.

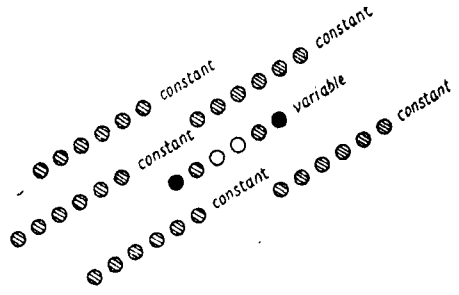


Fig. 6. Use of photographic plate, placed slightly beyond focal plane of telescope to discover eclipsing binaries.

THE FREQUENCY OF BINARY STARS

The question now arises as to the frequency of binary stars in our Galaxy. We cannot answer the question unequivocally because we cannot hope to detect all possible cases of 'duplicity' even among those stars which can be observed individually; and there are parts of the Milky Way where the stars are too numerous to be studied individually at all. The best we can do, therefore, is to make an estimate, on the basis of the results already achieved. And even these are not exhaustive, because we cannot avoid, in any of the three classes we have discussed, the effect of observational selection introduced by our methods of detection.

In the case of the visual binaries, it is clear that we cannot set an arbitrary upper limit to the angular separation, and study individually and exhaustively every pair of stars falling within that limit. Inset 25 shows a section of the Milky Way which would occupy for the rest of his life any astronomer rash enough to undertake its study. It has for long been the custom to specify a set of limiting separations, such as those proposed by R. G. Aitken, who suggested 40″ for stars brighter than magnitude 2; 20″ at magnitude 4; 10″ at magnitude 6; and so on down to 1″ for stars fainter than the 12th magnitude. Where there was independent evidence, such as common proper motion, that two stars might be physically connected, these limits would, of course, be ignored, but in general they formed a good set of working rules, which were sufficiently justified by the following argument.

The apparent separation of two stars depends not only on the distance between them, but also on their distance from us. Thus, in Fig. 7 we have four pairs of stars whose distance apart is the same in each case, but which are at different distances from the earth E. The distance apart we shall assume to be near the limit beyond which the two stars will become independent; and we see that, as we go further away from the earth, the corresponding angular separation becomes progressively smaller. But we cannot confine our interest to those stars

whose distances have been measured, so we adopt as a rough, but in the first approximation quite reliable, criterion of distance, the apparent brightness of the principal star. Thus we arrive at a convention like the one quoted above, in which the maximum allowable separation decreases with the brightness of the stars.

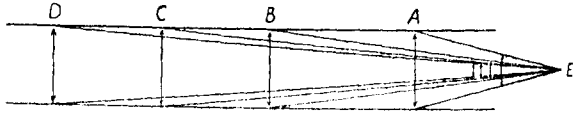


Fig. 7. Relation between angular separation and distance in binaries of the same actual separation.

It is clear, however, that any rule of this sort must introduce some degree of observational selection. It will lead, for example, to the ignoring of any pairs fainter than sixth magnitude, which happen to be so far apart and at the same time so near us that they are more than 10″ apart. And it has been claimed by some that this effect has given a quite false picture of the relative numbers of binaries among themselves, and in proportion to the total number of stars. However that may be, it is not disputed that Aitken's catalogue of 1932 contains a total of over 17,000 pairs, and extrapolation to allow for the part of the southern sky not included suggests well over 20,000, which is about one in eighteen of all the stars brighter than tenth magnitude.

In the case of the spectroscopic binaries, the features which influence the number of discoveries are the mass, the shape of the orbit, and its orientation. In general the orbital velocity increases in proportion to the mass of the system, and high velocities mean, of course, large displacements of the spectral lines. Thus, since large displacements are easier to see and to measure than smaller ones, the systems of large mass have a better chance of being discovered.

Again, the orbital velocity is influenced by the shape of the orbit. In a circular orbit the velocity is constant and, of course,

it may be small. But in an elliptical orbit the velocity is higher when the two stars are near each other than it is when they are widely separated. Thus the velocity will be high, as also will the possibility of discovery, at one part at least of the orbit. And, finally, since the Doppler effect is produced by velocities in the line of sight, it is clear that systems whose orbits are nearly edge-on to us will have a better chance of discovery than those whose orbital planes are more highly inclined to the line of sight. In the extreme case, a pair whose orbit is exactly perpendicular to the line of sight will have no velocity-component towards or away from us, and hence no Doppler effect, and its duplicity will never be discovered at all. There are nevertheless over a thousand known spectroscopic binaries, including some 500 whose orbits have been determined.

It is obvious that the eclipsing binaries will be subject to the same condition respecting the inclination of the orbit as the spectroscopic ones; an eclipse can take place only when the orbit is nearly edge-on to us. Then, too, we can more readily discover those pairs in which the surface brightnesses of the components differ considerably from each other. The eclipse of the brighter component will then produce a deep and easily noticed minimum in the light-curve. And finally, it is easier to discover the pairs with short periods, since it takes less time to observe and plot the variation of the composite light of the pair. It is clear again that we cannot hope to discover all the existing binaries by the eclipse method; but this method has given us some 2,000 pairs.

When allowance is made for all the double stars which, though accessible to the methods of observation used, have not yet been discovered, and for those which cannot be observed for the reasons discussed, it seems likely that at least a fifth of the stars in our Galaxy are double or multiple. Some investigators have suggested that the proportion may be as high as a half, and it seems that this higher fraction is found among the stars of mass greater than that of the sun. In any case, whatever the final verdict may be, the ratio of double to single stars is high enough to show that we are dealing with a

phenomenon of considerable importance, and one which must be taken into account in any theory of the origin and evolution of the stars. This question is one which is right on the advancing frontier of astrophysics to-day, and before discussing it we shall describe first some of the better-established deductions from the study of binaries.

DEDUCTIONS FROM THE STUDY OF BINARIES

The most important of these is undoubtedly the fact that double stars have enabled us to estimate the masses of the stars. There is, of course, no way of directly measuring the mass of a star. In the case of the earth, we can determine its mass by comparing its gravitational pull on some body with that of a known mass of lead, say. But in the case of bodies other than the earth, we must proceed indirectly, by means of Kepler's third law, and Newton's law of gravitation.

Newton showed that for two bodies of masses m_1 and m_2, revolving around each other in period P at a mean distance a, $G(m_1 + m_2) = 4\pi^2 a^3/P^2$. This equation may be taken to apply to a binary star. Similarly, for the sun and the earth, of masses S and E, for which $a = 1$ astronomical unit, and $P =$ 1 year, we have $G(S+E) = 4\pi^2$. Hence $(m_1 + m_2)/(S+E) = a^3/P^2$, where a and P are measured in astronomical units and years respectively. But if A is the semi-major axis of the orbit in seconds of arc, and p is the parallax of the pair, then $a = A/p$, so that finally, taking the mass of the sun as our unit, and neglecting that of the earth which is relatively very small, we have $m_1 + m_2 = A^3/p^3P^2$, where every quantity on the right-hand side can be measured, not for every binary system, but for a fair number. The relation, therefore, makes it possible to estimate the combined mass of the stars in a double.

Where the orbits of the two components separately about the centre of gravity are known, we can also obtain the ratio of the two masses, since these are inversely proportional to the semi-major axes. Then, knowing both $(m_1 + m_2)$ and m_1/m_2 we can find the individual masses of the two stars. In this way, the individual masses of a fair number of stars, and the combined masses of many pairs, have been found. And since the parallax, and hence the distance, is also known in every case, so also is the true brightness, or absolute magnitude.

From the data available at the time, the late Sir Arthur Eddington showed, in the early nineteen-twenties, that there exists a quite definite relationship between the mass and the absolute magnitude of a star. This relationship, which is shown in Fig. 8, shows that the luminosity of a star increases with its

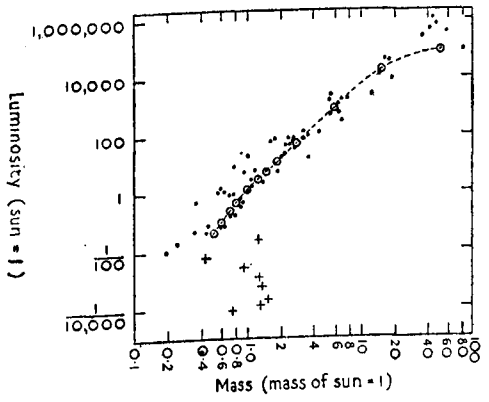


Fig. 8. Mass-luminosity relationship from McCrea, W. H. *Physics of the Sun and Stars*, Hutchinson (by courtesy of the author and publishers).

mass, so that, for example, a star of half the mass of the sun will be about a sixtieth as bright as the sun, while one four times as massive as the sun is about 100 times as bright. The mass-luminosity diagram, and the mathematical laws derived from it, are, however, more commonly used in the opposite direction, to tell us the mass of a star of given luminosity. The information derived in this way has been of the greatest importance in the building up of the present-day picture of a star; in particular, it has shown that the masses of the stars vary only from about a tenth of the solar mass to about 50

times that mass, while the total radiation emitted increases about 500,000 times.

This last piece of information leads to a picture in which the central temperature of the stars varies much less than their emission, so that the rate of generation of energy in the interior varies as about the 15th power of the temperature – a clue to the nuclear processes now believed to be responsible for the radiation of most stars. It may be too much to say that without double stars this insight into the structure and the working of the stars would never have been gained; but it is true that very much of modern astrophysics centres around the mass-luminosity relation. And that relation is based almost entirely on the double stars. The only other method of estimating stellar masses is from the gravitational shift of the lines of the star's spectrum. But this method is valid only for the white dwarf stars, which give us very little information about the conditions in a normal star.

Once the masses of a reasonable number of the binaries had been found, it became possible to use the equation $m_1+m_2=A^3/p^3P^2$ in reverse, and to derive the so-called dynamical parallaxes of many stars. This is possible chiefly because the combined masses of the components of most binaries do not differ very much from twice the masses of the sun. We may, therefore, as a first approximation, take $m_1+m_2=2$, so that $p^3=A^3/2P^2$, which gives a statistically useful estimate of the parallax, or distance. It may be noted, too, that since we have only the cube root of m_1+m_2 on the right-hand side, any given error in this term produces a comparatively small error in p.

It is also possible, where a pair can be studied as both a spectroscopic and an eclipsing binary, to calculate the radii of its components. From the spectroscopic observations, we can derive the orbital velocity of, say, the fainter star; and from the light curve we deduce the time it takes to pass in front of its companion. The circumstances of the eclipse are shown in Fig. 9, where we see that AB represents the diameter of the eclipsed star and AC that of the eclipsing one. Since we know

the velocity, and the corresponding times, the diameter may be calculated. It need hardly be said that, in practice, the actual work is often much more complicated, since, for example, the eclipse is not always central, and the light curves always show some observational irregularities.

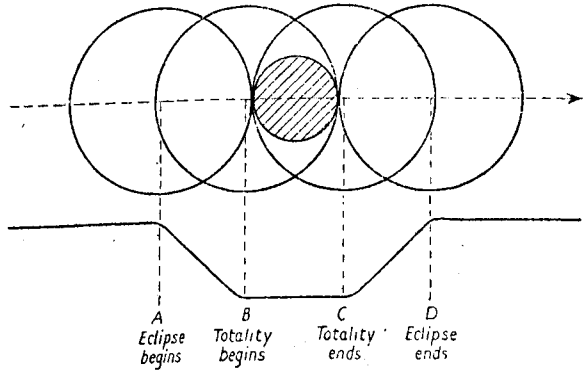


Fig. 9. Calculation of radii of component stars in pairs which can be studied both as eclipsing and as spectroscopic binaries.

The importance of these estimates of stellar radii is that they add to the small number of interferometer measurements available to check the radii calculated from the theory of radiation. The fact that these more or less direct measures agree with the calculated values indicates that the theory is sound.

The double stars have also contributed to the theory of relativity, and in a very important fashion. One of the fundamental postulates of relativity is that the velocity of light is the same in all circumstances, even when the source and the observer are in relative motion. This, of course, contradicts our normal experience of mechanical systems in which the velocities would normally be added or subtracted. Any direct evidence in favour of the relativity postulate is therefore important for the whole theory; and the orbits of double stars provide such evidence. The effect of any variation in the velocity of the light

from the star as it moved in its orbit would be to introduce an additional eccentricity into the orbit, and there would therefore be no zero or even very small eccentricities. Thus the circular and near-circular orbits which are known to exist prove that the velocity of light is in fact independent of any movement of its source.

All the investigations described so far are well-established and generally accepted results of the study of double stars. We come now to the more recent and more controversial matter referred to at the beginning, as being likely to influence very deeply our whole idea of the evolution of the stars. Since double stars are so plentiful in the Galaxy, it is clear that they are no casual and trivial phenomenon, but something fundamental in the stellar universe. Any theory of the origin and development of stars must therefore take account of them.

THEORIES OF THE ORIGIN OF BINARIES

At the moment, there are two main bodies of thought, which we may for convenience describe as those of Struve and of Hoyle. Each starts from the notion that stars condense from the matter scattered thinly through interstellar space, and they are agreed that very few binaries can have been formed by simultaneous condensation; but the two theories give very different accounts of the origin of the binaries from the stars so formed. In Struve's theory, the original condensation from the interstellar medium produces a rapidly spinning star of large mass and radius. As this star condenses further, it spins faster and faster, until its rotation becomes too fast for its gravitational cohesion. This rotational instability results in a splitting of the star, and the shedding of a ring of matter which may form a common atmosphere about the pair. As matter continues to pass from the stars to the gaseous envelope, whence it is dissipated into space, the nature of the pair changes. The size, mass, and rotational velocity of each component decrease, and their separation increases, until we are left with a pair in which one component has a gaseous ring around it.

So far, the process seems reasonable enough, and there are

many systems known which fit the theory quite well. The further course of the evolution of the binary is, however, rather uncertain. If the conditions are favourable, the double star may now become a single star accompanied by a planetary system. But there seems to be no certain guide to the way in which this will happen. One of the two stars in the system may move off into space, leaving one with a ring of gas, from which the planets may be formed in the manner envisaged by von Weizsäcker. But we are given no plausible reason why the binary should 'dissolve', nor any mechanism by which it may do so. Alternatively, the two stars may approach each other as matter is dissipated from the gaseous ring, carrying off both mass and energy of rotation. In this case, they will ultimately coalesce to form again a single star with a gaseous ring.

Perhaps the greatest weakness of the Struve theory is that it does not account for the widely separated pairs. These seem more likely to be accounted for by Hoyle's theory. Here again we start with single stars condensing from the interstellar medium. This time, however, they are thought of as being comparatively small and growing to giant size only in the cases where conditions are such that they can pick up matter from the surrounding medium. Double stars are then formed, not by the evolution of the single stars, but by their accidental association. Since the distances between single stars are normally very large, the majority of the doubles so formed will be widely separated. Then, if a pair becomes embedded in an interstellar cloud of low relative velocity, the accretion process will lead to an increase in the mass of each component. The two stars will then come nearer and nearer to each other.

This theory is almost directly opposite to Struve's; but it manages also to give an account of the origin of planets. In this case it is assumed that one of the two stars of a binary becomes a supernova, i.e., explodes with extreme violence. The explosion is assumed to be unsymmetrical, so that it pushes the rest of the star off, like a super-spaceship, effectively to infinity. At the high temperature produced, heavy elements are synthesized from hydrogen, and some of this new kind of

matter is captured by the sun, and in course of time the planets condense from it.

So far as concerns the origin of planetary systems, both of these double-star theories share the only feature common to all others: they have more opponents than supporters, and it cannot be said that the study of double stars has yet reduced the need for that activity.

As to the more important aspect of the two theories, it cannot be said either that they have given an acceptable picture of the evolution of double stars themselves. We have already indicated some of the weaknesses of Struve's theory. The weakest part of Hoyle's theory is the picture of double stars being formed by close encounters of single stars. It has long been believed that such encounters cannot be very frequent, and Struve quotes a result of V. Ambarzumian, that such close encounters as do take place are much more likely to separate existing pairs than to make new ones.

It is, however, not of vital importance that either of these two theories should be correct. What is important is that they have provided new views of the evolution of the stars, single as well as double. And whatever picture of the origin and evolution of the stellar universe is ultimately accepted, it will certainly owe something to the study of double stars.

FURTHER READING

As general references, neither of them easy reading:

Aitken, R. G. *The Binary Stars.* 2nd edn. New York, McGraw-Hill, 1935.

Struve, O. *Stellar Evolution.* Princeton University Press, 1950.

For theories of the origin of planets:

Jones, Sir Harold Spencer. 'The Origin of the Solar System'. *Science News,* **14,** 74 (1949).

Smart, W. M. *The Origin of the Earth.* Cambridge University Press, 1951.

Hoyle, F. *The Nature of the Universe.* Oxford, Blackwell, 1950.

For general stellar structure and the mass-luminosity law:

McCrea, W. H. *Physics of the Sun and Stars.* London, Hutchinson, 1950.

THE PERIODIC SYSTEM AND THE REDUCTION OF CHEMISTRY TO PHYSICS

H. R. PANETH

As the theory of a science advances, its laws become fewer but of greater scope. On that score, the biggest step in the theory of chemistry was the discovery of the periodic system which affords a natural classification of all the elements. The scope of the rules implied in the periodic system is illustrated by the fact that somebody who had never before heard of the element magnesium could tell, merely by looking at its place in the periodic table (Mg in Table I) that it is a metal (since it is near the top-left corner), forms a chloride of the form $MgCl_2$ (since it is in column 2), reacts with water about as violently as does lithium, and many other facts.

These rules, to which we appeal in 'predicting' the properties of an element from its place in the periodic table, are based on empirical regularities. There are exceptions to these regularities, and the rules have not yet been reduced completely to exact physical laws. Thus, while already in 1929 P. A. M. Dirac was able to say that 'the underlying physical laws necessary for a mathematical theory of the whole of chemistry are completely known', nevertheless textbooks of chemistry still run to a thousand pages without being exhaustive in the sense of a textbook on classical electromagnetism. This continued existence of chemistry as a collection of empirical statements is essentially due to two reasons:

First, these empirical statements themselves are sometimes about properties (such as 'characteristic smell') which have not yet been reduced to measurable quantities, or involve incompletely investigated conditions (the presence of traces of

moisture), or ill-defined states (substances apparently in equilibrium may actually be reacting very slowly). Thus one step in 'explaining' chemistry is the translation of rough terms into well-defined quantities. However, there remains a second difficulty arising from limitations in the practical use of the *theory* to cover even well-defined experimental results. Continuing the quotation from Dirac, 'the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble'. This is due to the fact that all chemical problems, even on the atomic scale, such as the interaction of two atoms, involve the interaction of a large number of electrons: they are 'many-body problems'.*

The exact calculation of any such many-body problem in quantum mechanics (as in classical mechanics) is so hopeless that there still exists in general a wide gap between the empirical facts of chemistry on the one side, and the fundamental laws of physics on the other. This gap is being bridged from both sides: on the one hand, there are some general consequences of the fundamental physical laws (for instance concerning symmetry) which serve as guides in approximate calculations starting from 'the physics end', and on the other hand there are some general rules in chemistry, arrived at empirically, which simplify the task of finally linking up all chemical data with the physical laws.

This process of closing the gap is the history of the periodic system. We shall first discuss the value of the system as an empirical rule, as a table implying a host of properties of elements, and its apparent limitations. We shall then see how far the periodic system can be derived from the fundamental laws of physics. Ultimately we will find that the best use of the periodic table, for predicting or explaining properties of elements, depends on empirical rules aided and amended by theoretical insight. This is illustrated by the case of the new elements.

*The classical example of a many-body problem is the exact calculation of the motion of the planets in the solar system.

THE CHEMICAL THEORY

There were two necessary preliminaries to the discovery of the periodic system: one, the discovery of *elements*, basic to the whole of chemistry; two, the discovery and determination of atomic weights, as distinct from equivalent weights.* Equivalent weights are simply measures of the characteristic proportions in which given elements combine, and are determined directly by ordinary chemical reaction and quantitative analysis, while the determination of the atomic weight requires in addition a knowledge of the number of atoms of the element in a compound molecule, which is inferred indirectly from molecular weight determinations and other physical measurements. This number essentially depends on the *valencies* of the elements involved. The valency of an element is defined as the number of hydrogen atoms (or their equivalents) that one atom of the element can bind. It is the most significant basic fact for the theory of chemical binding that one atom can interact chemically only with a limited number of other atoms. Sometimes an atom may have more than one valency, but in each case the valency is in general represented by a definite whole number.

The principle embodied in the periodic system is this: if we arrange the elements in an order which, with very few and easily explained exceptions, is the order of increasing atomic weights, we find that the valency changes smoothly from one element to the next in unit steps, increasing and decreasing periodically, and that the other properties of the elements change in the same periodic way. Thus there is a striking periodic recurrence of similar elements at certain intervals in this natural order of elements. This fact can be indicated by writing the sequence of elements in horizontal rows of length equal to those periods, so that similar elements ('homologues') fall below each other in vertical columns. The elements in any column have a common characteristic valency and are similar

*Sometimes called 'combining weights'.

in other respects. Table I is an example of such an arrangement.

The principle was first realized by Béguyer de Chancourtois in 1862, but the first clear publication, in great detail, was by Mendeléeff in 1869. We shall consider, below, the periodic table on the basis of our present knowledge; it was, however, one of the merits of Mendeléeff that he succeeded at a time when there were still several undiscovered elements, requiring 'gaps' in his table (which he pointed out as such).

We now know an unambiguous criterion for determining the 'atomic number', i.e. the place of an element in the natural order, namely the nuclear charge Z* (which can be determined from spectroscopic or still more direct measurements) and we are in a position to declare the periodic system 'filled' up to $Z=98$, and to confirm the chemical intuition which required in a few instances (e.g. A, argon, and K, potassium) a re-ordering of elements, not according to crude chemical atomic weights.

This is by no means the only example of far-fetched, detailed chemical evidence leading chemists to powerful generalizations ultimately culminating in and explained by physical principles (in this case atomic structure). Arguments based on such miscellaneous and entangled chemical data necessarily require greater intuition and ingenuity to reach such great generality than do those based on simpler physical data.

It is historically interesting that Mendeléeff dissociated himself from inferring solely from his scheme that there was one atomic material common to all elements.† Even though his

*Z=the number of protons in the nucleus: see Little, J. 'Models of the Atomic Nucleus'. *Science News*, **23**, 27 (1952).

†His discussion (*Ann. d. Chemie and Pharmacie*, Supplement VIII (171), p. 133 ff) on this point includes the suggestion (as an argument which would support the inference) that the failure of Prout's hypothesis (that atomic weights are simple multiples of a smallest weight) might be due to a mass-energy equivalence, only energy being conserved. This is a remarkable foreshadowing of the phenomenon of packing fractions and its explanation by the theory of relativity.

caution was misplaced, as events showed, we must compare this favourably with the inverse practice of starting from the assumption of simple numerology, and 'proving' it by finding some numbers to fit the data.

The merit of a theory lies not necessarily in 'prediction' in the temporal sense. That would depend on chronological order which may be fortuitous. It lies in the automatic inclusion of data, which may already be known, but which have not been put explicitly into the theory in the first place. We must judge any theory by the amount of data it covers *over and above* those it had to be fitted to.*

Thus there are some exceptions in the order of atomic weights in the periodic system. But they are so few as to make extremely satisfying the large number of chemical facts explained by the classification adopted. This classification depends not only on the sequence of the elements, but also on the postulation of six periods, which we can now take to be 2, 8, 8, 16, 16 and 32 elements in length respectively.

The extent of agreement is impressive even if we confine ourselves to properties known at the time of Mendeléeff's first publication. However, Mendeléeff's outstanding success was the prediction of the properties of three elements not yet discovered at the time. For the element required to fill a gap below silicon he was able to predict correctly not only the atomic weight and valency, but also the density of the element, its readiness to enter various chemical reactions, the physical properties of the resulting compounds, the acidity of the oxide, and the solubility of the hydrate. He was able to do this without any hypothesis regarding the structure of atoms, solely by interpolating between the properties of neighbouring elements in the periodic table, fitting the unknown element between the known ones in the manner of a crossword puzzle, to preserve a steady trend in the vertical as well as in the horizontal row.

*A figure-of-merit of a theory would be given by something like the total 'information' logically derivable from it minus the (linguistic) 'information' (defined in communication theory) involved in stating it.

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

Period	Group 1	2											3	4	5	6	7	0
I																	1 H 1·0080	2 He 4·003
II	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
III	11 Na 22·997	12 Mg 24·32											13 Al 26·97	14 Si 28·06	15 P 30·98	16 S 32·066	17 Cl 35·457	18 A 39·944
IV	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
V	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
VI	55 Cs 132·91	56 Ba 137·36	57–71 Rare Earths*	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
VII	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						

* Rare earths

VI 57–71	57 La 138·92	58 Ce 140·13	59 Pr 140·92	60 Nd 144·27	61 Pm 147	62 Sm 150·43	63 Eu 152·0	64 Gd 156·9	65 Tb 159·2	66 Dy 162·46	67 Ho 164·90	68 Er 167·2	69 Tm 169·4	70 Yb 173·04	71 Lu 174·99

The numbers to the left of the symbols of the chemical elements are atomic numbers, denoted in the text by Z. The numbers below the symbols are atomic weights relative to oxygen=16. For some unstable elements, the approximate weight of an isotope is given.

Such purely empirical rules concerning the periodic table are still useful in summarizing the general data of inorganic chemistry. Some of these rules are given in the next section.

EMPIRICAL GENERAL RULES

Table I is a convenient representation of the periodic system complete up to element 98 (californium). On the extreme right-hand side there is the column of the 'inert gases' He, Ne, A, Kr, Xe, Rn – helium, neon, argon, krypton, xenon and radon. They are unique in that they do not combine with any substances at all to form chemical compounds. It is thus reasonable to assign to them the valency zero, and to use them as a reference-line in considering other groups of elements. They are all momatomic gases of very low boiling points.

They are immediately followed in atomic number by Li, Na, K, Rb, Cs – lithium, sodium, potassium, rubidium, and caesium – and the recently discovered element, francium, respectively. These, the 'alkali metals', are 1-valent soft metals; they are very electropositive*, and increasingly so with increasing atomic number, reacting violently with water. They form strongly alkaline hydroxides, and increasingly soluble carbonates, while only a few compounds (such as the perchlorates) of the heavier alkali metals are insoluble. It is thus satisfactory to find that francium (whose atomic number, 87, is known from the nuclear processes producing it) also forms a soluble hydroxide and carbonate, but is precipitated together with Cs in the form of perchlorate.

The next column (2) in the periodic table is the group of alkaline earths, Be, Mg, Ca, Sr, Ba, Ra – beryllium, magnesium, calcium, strontium, barium and radium. These are again all metals, their positive electro-chemical potential increasing with atomic number, but are less electro-positive than their respective alkali neighbours. They are all 2-valent.

The inert gases are preceded in atomic number by the halogens, F, Cl, Br, and I – fluorine, chlorine, bromine, and iodine

*This term is in common use in chemistry, but in later discussion will be replaced by one more strictly defined.

– leaving on one side for the moment the recently discovered element, astatine (At). The halogens are again predominantly 1-valent (though iodine, for example, can also form the 7-valent IF_7), but very electronegative, though decreasingly so with increasing atomic number. They all form gaseous hydrides (in common with all elements up to four places before an inert gas in the periodic table), and these become acids in water, hydrogen iodide the heaviest; being the weakest acid.

The recently discovered element, astatine (whose atomic number, 85, again follows from the nuclear processes producing it), at first sight does not fit into the scheme at all well. True, it forms an insoluble compound with silver similar to the silver salts of Cl, Br, and I, and the element is dissolved to some extent in carbon tetrachloride, like iodine. But on the other hand, it behaves in many respects like a metal: it can be electrolytically deposited on the cathode as well as on the anode, depending on conditions, and forms a salt-like sulphide like a metal. In some respects, it is similar to its horizontal neighbour of the previous group, the metal polonium. This, however, is not entirely surprising, as the heaviest members of the groups in general show some 'horizontal' similarity; polonium in turn is rather similar to its neighbour, bismuth. And metallic properties, which already show slightly in iodine, generally become more pronounced with increasing atomic number in any group.

The same rule is illustrated in the preceding group, O, S, Se, Te, Po – oxygen, sulphur, selenium, tellurium, and polonium. Here oxygen is a non-metal, while tellurium is already a metal in most respects. The elements in this group are 2-valent with hydrogen, and up to 6-valent with oxygen and fluorine.

The group preceding this one, N, P, As, Sb, Bi – nitrogen, phosphorus, arsenic, antimony, and bismuth – is 3- and 5-valent.*

*With the exception of the first member nitrogen: the usual theory of valency does not consider N in N_2O_5 to be 5-valent. Part of the bonding is here by co-ordination (this term is explained, for instance, in Nyholm, R. S., *Science News*, **20**, 12).

The next group to the left, C, Si, Ge, Sn, Pb – carbon, silicon, germanium, tin, and lead – is 4-valent, and the one to the left again is 3-valent. Its members are B, Al, Ga, In, and Th; that is, boron, aluminium, gallium, indium, and thallium. Here the metallic properties already show in the second member of the group.

We can sum up in two rules:

(1) The valency of an element in combination with hydrogen and most other elements is equal to the number of places it is removed from the nearest inert gas in the periodic system. A maximum valency (with fluorine), equal to 8 minus that valency, is shown in many cases by elements 'on the right', that is preceding an inert gas.

(2) Elements are the more electro-positive (metallic) the farther they are to the left, so that the ones just beyond an inert gas are the most electro-positive, or the farther down they are in the periodic table. Conversely, they are the more electro-negative (the hydride being a strong acid in water), the nearer they are to the top right-hand corner. This is known as the diagonal rule. The most electro-positive element known as such is caesium (Cs); the most electro-negative is fluorine (F).

Transition Series and the Rare Earths

So far we have ignored two complications in the later periods of the table: first, the occurrence of the 'transition series' after the alkaline earths in periods IV, V, and VI; and, secondly, the further insertion of the 'rare earths series' at the beginning of the last transition period.

Each transition series consists of ten successive elements which show a good deal of 'horizontal similarity'. They are all metals; and, except the first and last elements and silver, they all exhibit several, in some cases a whole range of, valencies and formed coloured ions in some valency states. The first transition series includes some well-known examples: the green nickel, and the blue copper, salt-solutions.

While most transition elements have a 2-valent state, there is superimposed on this horizontal similarity a variation

analogous to the variation along the main periods considered above: thus scandium (Sc) is 3-valent, titanium (Ti) can be 4-, and vanadium (V) is 5-valent. There is a break at nickel (Ni), which resembles to a slight extent the break at an inert gas, and then 1-valent copper; while zinc has only a 2-valent state and marks the end of the transition series.

The two subsequent transition series are roughly analogous, and there is considerable similarity of elements in the same vertical column. But the diagonal rule, in particular, does not hold for transition elements; it is inverted, in the vertical direction in the table, as shown by the sequence Cu, Ag, Au – copper, silver, and gold – of increasingly 'noble' metals.

Element 43, technetium, which has recently been prepared artificially, illustrates the vertical similarity. It is more similar to Re (rhenium) than to Mn (manganese), being precipitated in analytical group II together with Re.

Finally, there is the rare earth series, a sequence of fourteen exceedingly similar elements, starting with $_{58}$Ce and ending with $_{71}$Lu, all very similar to $_{57}$La (lanthanum), the first transition element in that main period.* They occur in 2- and 4-, but mainly in 3-valent forms, most of them as coloured, paramagnetic,† positive ions. These properties are even more pronounced here than in the case of the transition elements.

TRANSLATION INTO PHYSICAL TERMS

In order to approach the explanation of these empirical rules by the fundamental laws of physics, we shall first reduce the many chemical properties referred to in rule (2) to a few more fundamental, well-defined physical magnitudes.

The most important physical magnitude determining the chemical behaviour of an element is its ionization potential. This is a measure of the energy required to remove one or more electrons from a neutral atom of the element. The smaller the

*In chemistry, these fourteen additional elements are lumped together with the preceding element, lanthanum, and the fifteen elements called 'rare earths'.

† See Nyholm, R. S., *Science News* **20**, 27.

ionization potential, the greater is the tendency of the neutral atom to become a positive ion (electrovalency) by losing an electron, and the greater the ionization potential the less is this tendency. The ionization potential is, in fact, a natural (but inverse) measure of 'electro-positivity', a term that we have been using without quantitative definition to characterize the extent to which an element is metallic, its tendency to react violently with water to form bases, etc.* Conversely, the ionization potential of an element may run parallel with the tendency of the neutral atom to acquire an electron and become a negative ion, although, strictly, this should be measured by the electron-affinity, the energy gained in absorbing an additional electron, which can be measured directly in some cases. This affinity should also be a measure of the tendency to acquire an electron by sharing with another atom (covalency), that is, of the tendency to form covalent compounds – another property which is implied in the rough term electro-negativity. The chemical properties summarized in these two terms, electro-positivity and electro-negativity, are reflected well in the trend of the ionization potentials. In chemistry, readiness to give up an electron, or to acquire one, is all.

We shall consider the curve of the first ionization potential, being the energy required to remove one electron completely from the neutral atom, as illuminating the general chemical character of an element (Fig. 1). The periodic variation of this quantity, as reflected in the diagonal rule, is clearly shown:

*Electro-positivity is not measured adequately either by electrochemical potential (the potential developed between a metal and a solution of its salts), or by the heat of reaction with oxygen. The first gives, for instance, the order Li, K, Rb, Na, instead of Li, Na, K, Rb. This is because an additional factor, energy of hydration, which depends inversely on ionic volume, enters into the electrochemical potential. Heat of reaction with oxygen is greatest for Li, and *decreases* in the order Na, K, Rb, Cs. Those familiar with thermodynamics will realize that even the particular affinity for oxygen is measured, not by the total heat-effect, but by the change of *free energy* accompanying the reaction.

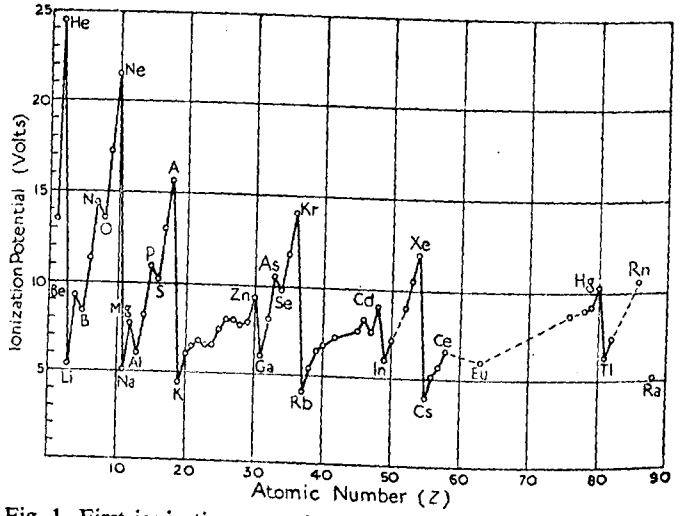


Fig. 1. First ionization potential (of neutral atoms), plotted against atomic number (redrawn from Herzberg, *Atomic Spectra and Atomic Structure*, Dover Publications, Inc., New York 19, N.Y., by courtesy of the author and publishers).

the energy required for ionization is a maximum for the inert gases and a minimum for the alkali metals. The height of the maximum decreases in successive periods, and this is true in general of corresponding points. The smaller energy required for ionization corresponds with the increasing electro-positivity at higher atomic numbers. The anomaly in the case of the transition elements is reflected less in the curve, by some corresponding points being higher in the later transition periods.

If we plot the atomic volume, i.e. the ratio of atomic weight to specific weight of the element in the solid state, against atomic number (Fig. 2), we find a similar periodicity, the maxima in this case occurring at the alkali metals, and the minima somewhere in the middle of the periods. The apparent atomic radius, defined for covalent compounds in such a way as to be largely independent of crystal structure, follows a similar curve.

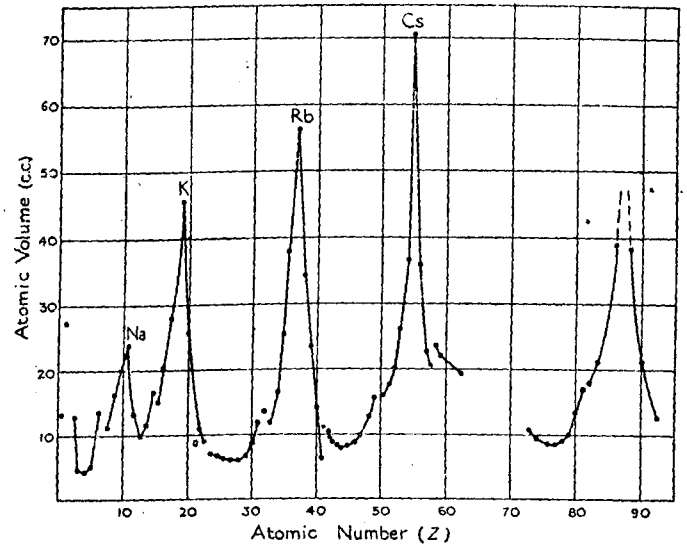


Fig. 2. Atomic volume (atomic weight divided by specific weight in solid state).

THEORY OF THE ATOM

In order to understand these facts theoretically, as far as possible, we have to bear in mind the wave-mechanical picture of the atom. This is essentially the Rutherford-Bohr model, consisting of planetary electrons revolving about a small, positively charged nucleus in orbits of various allowed radii, ellipticities, and orientations of orbital plane, but 'smeared out' by introducing the idea of a probability-distribution representing the distribution of a single electron in space.

We start with the hydrogen atom, consisting of one electron revolving about a nucleus which is a single proton. This is taken as a unit point-charge, attracting the electron by the well-known force of electrostatic attraction or repulsion called the Coulomb force. The attraction is balanced by the tendency

of the electron-cloud to spread out, to reduce the 'zero-energy'.*

This case of the hydrogen atom has been completely solved wave mechanically. The results of the calculation show that the only possible stationary states form a series, of which the successive members correspond to the different whole-number values of a main (first) quantum number $n=1, 2, 3 \ldots$ The corresponding states have a binding energy proportional to $1/n^2$. A lower quantum number n corresponds to a lower energy-state, meaning by this that more energy is required to lift an electron out of the atom, that is, the ionization energy, or binding energy, is greater.

These states correspond roughly to electron-orbits of radii proportional to n^2 in the Bohr-model approximation.

The wave-mechanical calculation shows also that for each first quantum number n, there are discrete states differing in a second quantum number l. This number measures the angular momentum of the electron's motion round the nucleus, and can range from $l=0$ to $l=(n-1)$, corresponding to orbits of decreasing ellipticity in the Bohr model: $l=0$ corresponds to maximum ellipticity, and close approach to the nucleus† for part of the orbit; $l=(n-1)$ to a circular orbit. Thus there are n different l-values for a given n. The states $l=0, 1, 2, 3, 4$ are referred to as s, p, d, f, g states respectively. An electron in a state such that $n=4$ $l=0$ (s state), is referred to as a $4s$ electron.

There is a third quantum number m, which, roughly speaking, determines the orientation of the plane of the electron orbit, and can have the integral values ranging from $-l$ to $+l$. Thus there are $2l+1$ possible m values for a given l.

Finally, there are two possible values associated with the orientation of the intrinsic electron-spin, and altogether there are $2n^2$ states‡ differing in one or more subsidiary quantum numbers but of the same main quantum number n.

*The so-called 'zero-energy', in quantum mechanics, is always associated with localizing a particle in a small space.

†The nucleus is situated at the focus of the ellipse.

‡Summing from $l=0$ to $l=(n-1)$: $2\sum(2l+1)=2n(n-1)+2n=2n^2$.

States of the same energy are known as 'degenerate' states. Since in the wave-mechanical solution of the problem of the hydrogen atom, the binding energy is found to depend on the first quantum number only, we have in the hydrogen atom $2n^2$ degenerate states of the same energy.

We now approach the structure of the lighter atoms – a problem which has not been solved exactly at all, being a many-body problem. The nucleus is treated as a point-charge of value equal to the atomic number times the proton charge. As a first approach, we shall consider a special case, in which an exact solution is possible. The effect of increasing the nuclear charge from the atomic number 1 (hydrogen) to a higher atomic number Z, *while keeping to a single outer electron*, is to increase the binding energy of each state by a factor Z^2. The degeneracy of each state corresponding to a main quantum number n remains $2n^2$ as before; the pattern of energy-levels has not changed, only the scale. This gives us, exactly, the relationship between the spectra of atomic hydrogen and, for instance, singly ionized helium, both having only a single electron; but the solution is too limited in scope to be of much use in the chemical problem.

Complications arise when we allow more than one electron in the atom. In that case it is not strictly true that we may describe the states of the electrons separately at all, since between them there is an interaction, mainly electrostatic, which leads to dynamic correlation.* However, we may in a good *approximation* consider the many-electron system of an atom as made up of electrons in separate, definite states; these states are no longer those corresponding to motion in the field of the nucleus only, but in that field modified by the presence

*In such a many-body problem, in the classical case, there is no equation for the orbit of any separate electron, but only an equation describing the system as a whole, owing to the dynamic correlation between the particles. This corresponds in quantum mechanics to the fact that strictly the separate electrons are not in well-defined states, but we can talk only about the state of the system as a whole.

of all the other electrons. This modification is described approximately by a static field superimposed on the Coulomb field of the nucleus, and representing the time-averaged effect of electron-electron interaction. This consists largely in a screening effect: the electron whose state we are considering is screened to some extent from the nucleus by other electrons. Electrons predominantly further away from the nucleus in their probability-distribution, i.e. corresponding to higher main quantum numbers, will have a negligible electrostatic effect on the electron in a given state, just as a uniformly charged hollow sphere exerts no electric force inside itself. Electrons predominantly nearer to the nucleus will have a screening effect. Thus we describe each electron separately as being in a state modified to allow for the influence of the others. We like to have such separate-electron models, since valency consists in the addition or removal of electrons.

The state in which we find an atom at normal temperatures is the state of lowest energy, or ground-state. We might be tempted, then, to predict the ground-state of any neutral higher atom as consisting of the corresponding number of electrons all in the lowest (most tightly bound) state of main quantum number $n=1$, each electron experiencing a binding energy equal to Z^2 times the binding energy in hydrogen, but modified slightly by interaction between the electrons whose probability-distributions completely overlap. This, however, is forbidden by the Pauli principle, which we shall next consider.

The Pauli Principle

This principle, which is basic to the theoretical understanding of the periodic system, is fundamental in the present state of physical theory; it has to be added to the other fundamental laws of physics, such as the uncertainty principle, for it cannot be derived from them. It is connected with the 'indistinguishability' of elementary particles. Two elementary particles of the same species, such as electrons, are absolutely indistinguishable from each other apart from being in different states (characterized, e.g. in the hydrogen atom, by the

four quantum numbers). In fact, we should not talk about 'individual' particles at all. The electron density is, in a sense, quantized, occurring in charge-packets which we call electrons, but there is absolutely no 'individuality' attached to them: two electrons are not merely indistinguishable from a practical point of view; they are indistinguishable *in principle*. And this fact leads to physical results, connected with statistics, which are quite different from the physical results which would arise from mere extreme similarity. Two electrons are just parts of a 'collective' world distribution of electron-matter, and exchange of one electron with another leads not merely to a practically indistinguishable distribution, but *is* in fact the same distribution. The physical state of the system is thus completely described by a given density distribution.

In quantum mechanics, the state of such a collective is completely described by a 'wave function'. Any physical quantity, such as the density of distribution, depends on the square of the wave function. Thus for any given density-distribution of the collective, only the absolute value of the wave function is fixed – that is, its value apart from sign – and we have still a choice between a factor of $+1$ or -1 for our wave function. Thus the wave function may change sign, or remain unaltered, when we 'exchange' two electrons, since the only physically observable quantity, its square, remains unaltered, as it must do for 'indistinguishability'. The former alternative, change of sign on exchange, is referred to as anti-symmetry, the latter as symmetry.

All particles of a given collective, that is a given species, enter the wave-function in the same way, either all symmetrically, or all anti-symmetrically. It is logically impossible that some electrons should be symmetric in the wave-function, and some be anti-symmetric, since inconsistencies would arise. Which of the two choices applies depends on the species: for electrons, it is found empirically to be anti-symmetry.*

It is clear that two electrons, entering anti-symmetrically into

*For some species of meson, the wave-function is symmetric; this statistical alternative is connected with the 'spin' of the particle.

the total wave-function, cannot both be in the same state. Thus at most two electrons, differing in spin, may have the same spatial distribution. For instance, in the case of the helium atom in its ground-state, we have two electrons in the $n=1$ state. This is the ground-state corresponding to the single-electron ground-state, modified slightly by the interaction between the electrons.

Building up Higher Atoms

If we could neglect electron-electron interaction, the pattern of energy-levels would be the same as that for hydrogen: we should have the first two electrons in the lowest level, $n=1$, eight (2×2^2) further ones at a considerably higher energy,* $n=2$, eighteen (2×3^2) in the next level, and so on.

The effect of electron-electron interaction is to change the sequence of levels, both for a given nuclear charge, as we place more electrons into the atom, and actually also for a given number of electrons as we go from one nuclear charge to another, with increasing ionization. (Starting from a neutral atom, the first type of operation would produce, successively, negative ions of the same element, singly charged by one added electron, doubly charged by two, etc. The second type of operation would produce positive ions: first, a singly charged ion of the next higher atomic number, then a doubly charged ion of the atomic number following that, and so on.) There are two main effects:

(1) The screening of the nucleus, as seen by an electron, by the intervening electrons reduces the nuclear charge to an effective atomic number roughly equal to $(A-C)$, where C is the number of intervening electrons. This means that the states of different n are still more widely separated, the higher n states being less tightly bound. This explains the sharp break in the periodic system at helium (He), which has a high ionization potential, and is completely inactive chemically, being unable either to lose an electron without great expenditure of

*The energy-gap would be Z^2 times the energy-gap between $n=1$ and $n=2$ for hydrogen.

energy or to bind an additional electron in a higher orbit. The next element lithium (Li) has one electron in the outer orbit, $n=2$; this experiences an effective nuclear charge roughly equal to $3-2=1$, and behaves very much like the single (valency) electron of hydrogen.

(2) Apart from this overall shift in the successive quantum levels due to screening, the degeneracy (equality of energy) of a given main quantum level is destroyed: the more highly elliptical orbits approach the nucleus more closely and are therefore affected less by screening than the circular orbits, $l=n-1$, which are practically screened from the nucleus by the full charge of the electrons of lower n.

Thus, in a many-electron atom, the energy of the highly elliptical $s(l=0)$ state of a given main quantum number, corresponding to an orbit approaching the naked nucleus, may well be lower than the energy of the $d(l=2)$ state, of main quantum number one less, since the d state corresponds to an orbit which is more nearly circular and largely screened from the nucleus. This is indeed found to be the case. If electron-electron interaction could be neglected, it would be expected, as already stated, that successive energy-levels, or 'shells', would be completed by successive additions of 2, 8, and 18 electrons; and, therefore, that we should find corresponding values of the nuclear charge Z for neutral atoms in which these shells have been completed (viz. the inert gases). In fact, we find the first electron-shell completed at $Z=2$ (helium), and the second shell at $Z=10$ (argon); but the $n=3$ shell is completed with argon at $Z=18$, not at $Z=28$. Actually, the big break is found after only the s and $p(l=1)$ states, of main quantum number 3, have been filled. The d states would be so well screened as to be at a much higher energy-level. After that, the next two electrons are found in the $4s$ state, that is in a state of next main quantum number, which brings us to calcium ($Z=20$). These assignments of states are confirmed spectroscopically.

In the next element, scandium (Sc), we find a $3d$ electron which here lies lower than the two $4s$ electrons. So in scandium

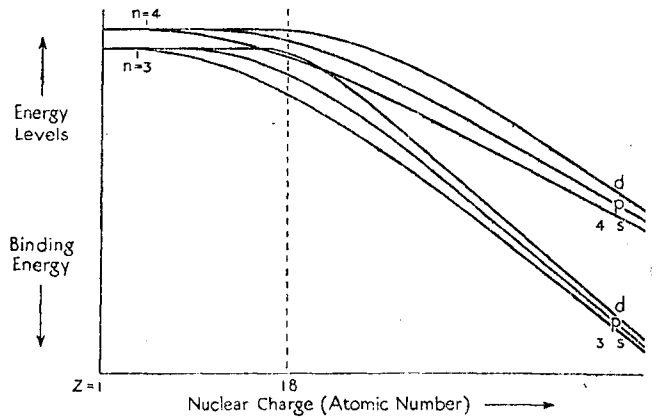


Fig. 3. The effects of the screening of electrons from the nucleus in neutral atoms of increasing charge Z (atomic number). Screening is by electrons predominantly nearer to the nucleus. Each group of curves (three in each group) represents the energy-levels of electron-states of the same main quantum number ($n=3$, and $n=4$, respectively), as the nuclear charge, and so the number of electrons, increases; the curves are not to any scale. With no screening, all curves would drop progressively (binding energy increasing) as nuclear charge increases, and the curves in each group would coincide, the main quantum number deciding the energy-level (as in hydrogen, $Z=1$). In fact, the fall in energy is delayed, and there are differences in energy between states of the same n. The s state (second quantum number, $l=0$) is the most highly elliptical and the least screened; in this, the fall in energy is least delayed. The fall is most delayed in the d state ($l=2$), the least elliptical and the most screened. At $Z=18$ (the inert gas, argon), the $3d$ curve lies higher than the $4s$ one; the order in increasing energy is $3s$, $3p$, $4s$, $3d$. Eight electrons go into the $3s$ and $3p$ states before the 4-shell is begun. Their number corresponds with the difference of 8 in atomic number between argon and the previous inert gas; this difference would be 18 (ten $3d$ electrons extra), if there were no screening. With further increase in nuclear charge, screening becomes less important relatively, and the $4s$ and $3d$ curves again cross (at $Z=21$, scandium).

the level-sequence found in calcium is actually changed on increasing the nuclear charge. This is reasonable, since we may sum up the matter as follows:

For low effective nuclear charge, the separation in energy of the s, p, d, etc. states, due to differences in screening, is important. As we increase the nuclear charge, a point is reached at which the effective nuclear charge is so large as to restore the level sequence in the order found in hydrogen (*see* Fig. 3).

Apart from such complications, we can use the 'building-up principle' as a rough guide; it states that the level-sequence in a given neutral atom is essentially the same as that in the singly ionized atom of the element immediately following, the doubly ionized atom of the element following that, etc.

GENERAL FEATURES OF THE PERIODIC SYSTEM

We are now in a position to explain the general features of the periodic system, which are illustrated by the curves already given showing two basic properties of the element, the first ionization potential and the atomic volume. Since the valency interaction of atoms consists essentially in the partial (covalency) or complete (electrovalency) transfer of electrons from one atom to another, the numerical valency, and the kind, and the strength of the valency-bond is essentially determined by the energy gained or lost by each participating atom in gaining or losing electrons. The tendency towards ion-formation (electrovalency) is essentially measured by the ionization potentials of the individual atoms, while the formation of covalent bonds is only described in a very rough approximation in terms of the quantum states of the separate atoms.*

The energy required to remove one electron from the neutral atom is a minimum for elements immediately following an inert gas: here there is one electron in an outer orbit, relatively easily removable to leave a 'core' of remaining electrons

*Compare article by R. S. Nyholm, already quoted. The different kinds of valencies are discussed clearly and more fully in Sidgwick's *Electronic Theory of Valency*.

distributed essentially as in the preceding inert gas. Any further removal of the electrons would require breaking up the inert-gas structure by removal of an electron from a lower shell against an attraction even greater than in the case of the inert gas (the nuclear charge being one unit greater). Thus the alkali metals are electropositive, readily *losing* the negative charge of one electron, and are 1-valent. The neighbouring alkaline earths have two electrons in the *s* state of the highest main quantum number, which again are relatively easily removed, but with a greater expenditure of energy, since here the effective nuclear charge has increased by almost one unit even for the first electron to be removed, and by exactly one unit for the second electron to be removed. Hence these 2-valent alkaline earths are less electropositive than their respective alkali neighbours.

As we go from one element to the next higher* in the same group, the atomic volume increases, corresponding to the increase in radius with n, and the removal of an outer electron becomes easier owing to the increased distance from the nucleus. Thus we expect increased electro-positivity in the higher periods.

On the other hand, the atom of a halogen, immediately preceding an inert gas, is able to accommodate one more electron without going to a higher quantum number. Indeed, energy is gained by so filling the state, the effective nuclear charge for the state in the neutral atom being by no means zero. It is, in fact, this energy-gain on the part of, say, chlorine (Cl) on acquiring one more electron to become a singly negatively-charged chlorine ion, that more than compensates for the small energy-expenditure required to ionize sodium (Na), thus explaining the stability of the Na^+Cl^- structure. The effective nuclear charge for sulphur, one place before chlorine, is of course less, and hence also its tendency to acquire additional electrons.

*By 'higher' element is meant an element of higher atomic number; the 'higher' elements and periods of elements are placed lower in the periodic table.

We have thus explained the diagonal rule, that elements are the more electro-positive the smaller the group-number and the higher the period-number. The tendency towards the formation of covalent bonds, being essentially another way of acquiring additional electrons, follows the same trend as the electro-negativity.*

There are, however, exceptions to the diagonal rule: the transition elements. These are elements which we have characterized earlier as exhibiting variable valency and forming coloured ions. We have also identified their first occurrence in the periodic system (with Sc) with the beginning of the filling-up of the sub-level of the 'inner' main quantum level. It has not been possible so far to calculate theoretically the exact atomic number at which the *d* sub-shell would be expected to begin. This question, being a many-body problem, can at present be solved only by successive approximation. The most successful method so far has been to treat the whole atom statistically, essentially as a quantized gas of mutually-repelling electrons in the field of the nucleus: this method, surprisingly, gives roughly correctly the points of first occurrence of the *d*-shells and of the *f*-shell ($l=3$) respectively.

While we are thus unable to predict the exact point of first occurrence of the *d*-shell entirely from theory, we can make the actual occurrence at Sc very plausible both qualitatively (see above) and quantitatively (by statistical methods). This semi-empirical situation is typical of the theory of the periodic system and, indeed, of the whole of chemistry.

We can predict quantitatively the exact end of the *d*-shell

*The number of electrons that an atom of an electronegative element will acquire by formation of covalent bonds is, in general, equal to the number required to complete an inert gas configuration (if we consider the separate atom). But in any case it is not greater than the number that can be accommodated in a shell of the main quantum number of the valency-shell. Thus there is a covalency maximum for a given period; the highest covalency found (as in OsF_8), corresponding to 16 shared electrons in the valency shell, can clearly only occur in the higher periods.

with considerable confidence. Once the d-shell begins to be filled (at Sc; $Z=21$), we would expect the process to continue as we increase the atomic number, rather than a filling of levels of higher n; this is because we are now 'over the hump' due to screening, and the increasing nuclear charge favours the hydrogen-like sequence with the d-shell of lower n being increasingly tightly bound. Now the Pauli principle tells us that there are exactly 10 vacancies in the d-shell, and so we would expect the 31st element, gallium (Ga), 10 places removed from scandium, to be again a homologue of aluminium (Al) in Group 3, having three electrons in the outermost level, as is indeed the case. Moreover, there is a break in the transition period at $Z=28$ (nickel), when the total number of electrons over and above the argon-like core ($Z=18$ for argon) is just sufficient to fill the inner d-shell. The next element, copper (Cu), exhibits 1-valency as well as the more common 2-valency.*

The general features of the transition elements are well understood. Their variable valency is due to the closeness of the energy of the (inner) d-level to the energy of the p-shell of the next higher main quantum level, allowing electrons to be transferred to outer (valency) orbits with an expenditure of energy smaller than that gained by the chemical combination. Secondly, the circular inner orbits of, for instance, the $n=3$ d-shell do not expose the electrons sufficiently to act as valency electrons, and hence the successive transition elements show horizontal similarities in states where they differ only in inner occupation-numbers. Thirdly, the colour of ions of transition elements in certain states is due to vacancies in the inner d-shell allowing transfer of electrons from one sub-state of that shell to another. We have explicitly neglected this kind of 'fine structure' in our chemical considerations above, because the energy-differences between such sub-shells are very small: they correspond to the absorption of light quanta in the visible range. It is, in fact, a test for vacancies in the sub-shell to look

*There is also a minor break at Hg, which formally has some similarity to the He electron-structure, and also forms monatomic vapour.

for coloured ions: 1-valent Cu is colourless, the inner d-shell being filled by 18 electrons; 2-valent Cu has a blue ion, there being a vacancy in the d-shell. We shall make use of such criteria in discussing the transuranic elements, those of $Z=93$ and upwards, which have been made artificially.

The higher valencies of transition elements will be exhibited in reactions with the more electro-negative partners. While the maximum valency is clearly limited by the number of electrons available in the outer shells and the next lower d-shell, and hence increases along the period, this high valency will become energetically less and less favourable,* owing to the increasing effective nuclear charge acting on the d-shell.

While the horizontal variation in properties of transition elements is generally understood, the vertical variation is not: an example is the decrease in electro-positivity, referred to above and partly accounted for by the anomalous variation in ionic volume. In particular, silver seems to be exceptional.

The first long period, having filled the inner d-shell and the outer s-shell at zinc, continues by filling up the p-shell of $n=4$; we again reach a break after the p-shell has been filled up at the inert gas krypton. After that, the s-shell of $n=5$ is filled at strontium, followed by the inner (now $n=4$) d-shell as before.

A new situation arises in the third long period, when, after lanthanum (La), a series of mutually very similar elements is found, the so-called rare earths. Here, clearly, a shell is being filled that does not allow free interchange with the outermost, valency, shell. Presumably the shell being filled is the hitherto vacant f-shell ($l=3$) of $n=4$. The beginning of this rare earth series occurs about where calculated on the statistical model.

We are again in a position to predict exactly where this series should end. The Pauli principle allows $2(2l+1)$, that is 14, f-electrons. Thus we expect the rare earths to end at element 71, and element 72 should be a homologue of zirconium, and no longer a rare earth. This point was of great value in the

*i.e. it will be the lowest energy state only in external conditions which are particularly favourable.

discovery of the element, which in the periodic table had long been conspicuous by its absence. Till the application of the Pauli principle by Bohr, it was expected that element 72 would be a rare earth, and it was looked for in these minerals. On the basis of the Pauli principle, Bohr predicted its similarity to zirconium, and indeed its x-ray spectrum was found in zirconium minerals shortly afterwards, and the new element, hafnium (Hf), separated by fractional crystallization. Here theory was necessary in addition to empirical rules.

Up to this point, then, we have explained the existence of the main periods of 2, 8, 8, 18, 18, and 32 elements. There remain for consideration the elements of highest atomic number.

TRANSURANIC ELEMENTS AND THE RECURRENCE OF THE RARE EARTHS

In considering where the next 'rare earths' series starts, we run across two difficulties. One is that the $6d$ and $5f$ levels lie very close to each other in energy; and because of this closeness we have a further uncertainty of definition: f-levels may be occupied in one state of an atom but not in another.

We would expect generally a tendency for more electrons to be 'pulled into' the valency-shell when the atom combines with others (lowering of energy by chemical binding). It is, in fact, found that the *chemical* properties of the elements near uranium are roughly those of transition elements, with valency increasing with atomic number, while in the isolated atom there are occupied $5f$ levels, revealed by spectroscopy or magnetic measurements. Again, f-levels tend to be occupied still more frequently in the ions than in the neutral atoms.

The chemical tendency to change more f-electrons into valency-electrons available for binding will decrease with increasing atomic number, as will the electro-positivity. Similarly, the maximum valency corresponding to the highest state of oxidation is found to drop in the later elements, indicating that we are well inside a 'rare-earths' series.

The general chemistry of the transuranic elements is best

illustrated by the case of plutonium (Pu; $Z=94$), about which most is known. In trying to guess its properties, we might be guided by analogy either to osmium, treating all elements in the last period from actinium (Ac; $Z=89$) onwards as transition elements, or to samarium, treating these elements as rare earths.

If the analogy is to osmium, we should expect in general higher valencies, up to 8, as in OsF_8. If the analogy is to the rare earths, we should expect mainly 3-valencies, and a close similarity between elements.

In the preceding elements, the analogy to the rare earths is certainly not perfect: thorium and uranium differ much more than do cerium (Ce) and neodymium (Nd), although thorium is very like cerium. Furthermore, uranium is most stable in the 6-valent state, and thorium in the 4-valent state. Thus they are more like transition elements than rare earths.

Passing now to plutonium, from 'the physics end' theory leads us to expect all plutonium ions to be coloured (since there are still vacancies in inner shells), and at least odd-valent compounds and ions to be paramagnetic (since these would have an odd unpaired electron). Merely by empirical extrapolation, we expect a tendency to the formation of complex ions. The element is presumably a hard, high-melting, metal.

In fact it is found that plutonium is most stable in the 4-valent state, but also occurs in the 3- and 6-valent states. This is in accordance with the general phenomenon that *rare earths* tend to shift their stable region to lower valencies with increasing atomic number, since the effective nuclear charge increases; while the *transition elements* may exhibit their maximum valencies, which increase steadily from one atomic number to the next, by shifting all available electrons from the d-shell (which compared with the f-shell is less deep-lying) to the outer shell. However, the analogy with the rare earths is certainly not complete yet, even in the region of plutonium, which differs considerably more from its neighbour, neptunium (Np; $Z=93$) and is much more readily separated from it than is the case with two neighbouring rare earths.

However, the two following elements, americium (Am; $Z=95$) and curium (Cm; $Z=96$), do show similarity and difficulty of separation comparable to the rare-earths case.

Thus the appearance of the second rare-earths series seems to be delayed in some respects. The highest elements become more rare-earth-like (just as they tend to have lower valencies) owing to increasing effective nuclear charge. But the valencies of the 'rare earths' are then 'as if' the rare-earths series started with thorium. (The first rare-earths series starts with the homologue of thorium, cerium.)

In other words, although the rare-earths period seems to start chemically only somewhere after uranium, when it does start it tends to have all but three electrons in the *f*-shell. This is shown by the 3-valency of all elements from uranium onwards up to and including berkelium (Bk; $Z=97$): starting with Am, 3- is the most stable valency state. It is also shown by an interesting analogy of curium to gadolinium, eighth in the rare-earths series, in forming a colourless 3-valent ion, indicating that there is no empty state in the *f*-shell close enough in energy to the ground-state for transition to it to involve the absorption of light quanta in the visible range, although in fact there is ultra-violet absorption, corresponding with a greater energy-difference. Again, curium has only a 3-valent form, while its predecessor, americium, shows 2-valency, as does europium, its homologue in the lower series.

The occurrence of a break after americium is explained by the following rule: 'other things' being equal, electrons tend to be in different spatial states as long as possible owing to their mutual repulsion. By 'other things', in this context, are meant first and second quantum numbers. Thus electrons will, if possible, avoid pairing into the same *m*-state. This ceases to be possible when we have more than 7 *f*-electrons – americium being the seventh element from radium – and hence there is a break at this point. Further, *f*-electrons can be recognized by their strong paramagnetic effect. On this evidence, it appears that the atoms, or ions, of transuranic elements have all their electrons beyond 86 (inert gas structure) in *f*-states. This again

is consistent with the impression of a rare-earth period, if with a delayed start.

The reason for the delayed start of the second rare-earth period is not understood, just as the variation in the successive transition periods remains to be explained. It is likely that the answers to such questions do not involve additional physical effects, but demand an improved computational approach.

CONCLUSION

The periodic table is a useful map of the main properties of the elements. Empirically we can predict those properties which vary smoothly from neighbour to neighbour (vertically or horizontally), while theory predicts best just the step-like variations, explained by the integral number laws. These integral numbers are a direct consequence of the existence of the electron, the quantization of states, and the Pauli principle.

HYDROELECTRIC AND IRRIGATION DEVELOPMENT IN THE U.S.S.R.

THE current plan of river utilization and development in the U.S.S.R. is illustrated in Inset 1–19. The scheme is unusual in respect of the large area covered, the boldness of its conception in geographical terms, and the many projects on which work is proceeding. It was begun some three years ago, and is due to be completed in 1957. Its scope can be seen from the panoramic map reproduced as Inset 14.

Inset 2–6, 8–10, and 17 illustrate a section of the scheme which is approximately complete: the Tsimlyanskaya hydroelectric station on the lower Don; the Volga–Don navigation canal, entered from the Don through the Tsimlyanskaya reservoir; and the Don Trunk Canal, fed from the reservoir, for the irrigation of some 1,800,000 acres between the Azov Sea and the northern end of the Caspian. Inset 17 shows a section of one of four tunnels, each 6 km. long, in which the irrigation canal is being carried beneath the ridge which forms the Don–Sal watershed. The hydroelectric station, with an installed capacity of 160,000 kilowatt, and the Volga–Don canal should have settled down to normal operation after the spring floods. On the irrigation side of the project, time will be needed for the reconditioning of the land irrigated and for bringing it into cultivation.

As the example illustrates, this is a multi-purpose scheme. It is at the same time a regional development plan for South Russia, based primarily on irrigation; a programme of hydroelectric development, notably on the Volga, but extending also to the Caucasus and the Aral–Caspian region; and a contribution to inland water-transport. The advantage of such an approach has been increasingly realized during the last 15–20 years, in the United States among other countries, and it has

ABOUT OUR CONTRIBUTORS

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

J. A. NELDER read mathematics at Cambridge, afterwards obtaining a Diploma in Mathematical Statistics. Aged 27, he is now statistician to the recently formed National Vegetable Research Station at Wellesbourne, Warwick.

H. R. PANETH studied chemistry at Oxford and physics at the University of Chicago. First worked in experimental physics, then in theory. Now teaching at Chelsea Polytechnic, London.

H. S. A. SHERRATT is a young graduate, in chemistry, of Birmingham University. He has done research in the Cancer Research Laboratories at the Birmingham Medical School, and is now working on plant proteins at the Department of Agricultural Chemistry, University College of North Wales, Bangor.

P. L. SHORT graduated first in philosophy and English at Bristol, then at Oriel College, Oxford, in psychology. Aged 27, he is now research psychologist at the Burden Neurological Institute, Bristol, working on psychological factors associated with electroencephalographic recordings.

D. L. WOODHOUSE graduated with 1st Class Honours in chemistry, Birmingham University. Ascough Scholar in Chemistry and Sir Charles Hyde Scholar for Research in Mental Disease, 1926–30. Fellow of the Royal Institute of Chemistry. Is now Officer-in-charge of Biochemical Laboratories of the Birmingham Branch of the British Empire Cancer Campaign; Senior Lecturer in Chemical Pathology at the Medical School; and Research Fellow in Cancer Research, University of Birmingham.

C. B. WILLIAMS is a Cambridge graduate, with Honours in Natural Science, a Diploma in Agriculture, and Sc.D. By profession a research entomologist, but specially interested in the application of statistics to biological problems, animal and human; also

been pointed out that 'it is possible to obtain certain benefits at less cost than ... with single-purpose projects, and to realize other benefits not otherwise economically attractive'.* A number of such projects have been constructed, and the Russian scheme is remarkable rather for the thoroughness with which combined development is being applied to an alluvial river system. (Compare, for example, the alluvial rivers of India and Pakistan, where the main effort has been in irrigation and flood control.)

In its regional aspect, the main area of the scheme extends from approximately the Ural River to the mouth of the Dnieper, some 900 miles from east to west, and from the region of the Don–Volga watershed nearly to the Caucasus foothills, some 500 miles from north to south. For the three main rivers, the order of the target-dates for completion is: Don, Volga, Dnieper. To these may be added the area in Turkmenistan and Uzbekistan watered by the Amu Darya River (the Oxus). The two areas are to be linked during the last stage of the scheme by the construction of the 700-mile Main Turkmen Canal. It is planned to divert some 60 per cent of the water of the Amu Darya River, which now flows into the Aral Sea, to flow instead through the Turkmen Canal into the Caspian; with a chain of hydroelectric stations down the canal (total head, some 300 ft), irrigation for 3,000,000 acres of waterless clay desert, and navigational access to the Aral Sea, subject to changes in level.

The project is to serve a further purpose in relation to the main scheme. Because the Volga flows into an inland sea, the Caspian, on which there are inland ports, notably Baku, effects on the level of the Caspian may be of long-term importance. Irrigation for 6,000,000 acres is to be provided by the Volga, as well as water for livestock over a further 25,000,000 acres. At the equivalent of 2 ft depth of water per year over the irrigated area proper, this alone would absorb just under 5 per cent of the mean annual flow of the Volga. In addition, there

*Wrather, W. E., and others. 'Energy Resources of the United States'. Fourth World Power Conference (1950).

will be the livestock areas, and losses due to increased evaporation from the reservoir lakes on the main river and from irrigation canals, so that the loss from the river-flow might amount to some 10 per cent. This is consistent with a further estimate that the increased evaporation resulting from *all* sections of the scheme (i.e. not only the Volga) will be equivalent to about 30 per cent of the mean flow of the Volga. Some part of the water thus evaporated might be expected to return to the river through precipitation within the catchment area. But the level in the Caspian has been falling in any case in recent years, and it is expected that there will be a deficit to make good. There have already been two minor additions to the Volga watershed through the diversion into its tributaries of the headwaters of the Pechora and Vychegda Rivers, of which the lower waters flow respectively into the Barents and White Seas. A more significant addition will result from the diversion from the Aral to the Caspian of 60 per cent of the flow of the Amu Darya River. This contribution should amount to 6–7 per cent of the Volga flow, and it will be an addition which could be controlled. The level in the Aral Sea has been tending to rise lately, but after the diversion would fall.

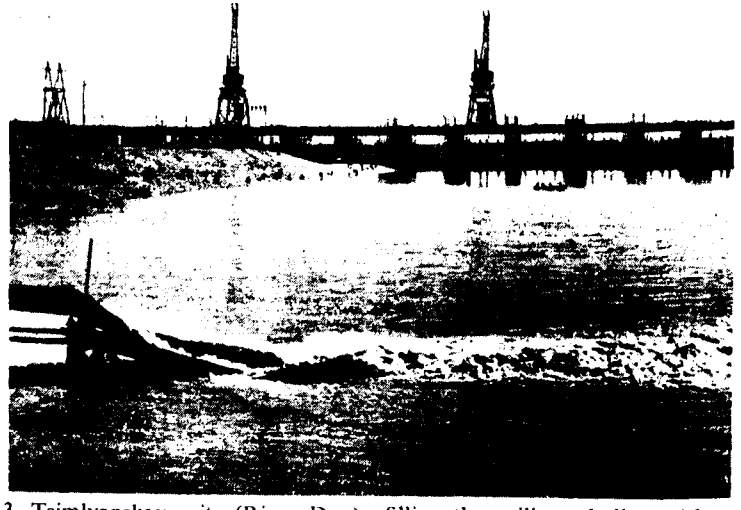
The plan as a whole provides for the irrigation of some 6,000,000 hectares (about 15,000,000 acres), and for the provision of water for livestock over a further 20,000,000 hectares (about 49,000,000 acres). For proportion, the total area irrigated in India and Pakistan is about 40,000,000 acres. The area of cultivable land scheduled for irrigation under the present Russian schemes is comparable with the irrigation-system built up in the Punjab over a considerable period. Changes in methods of field-irrigation have been adopted (Inset 15–16): the use of temporary (seasonal) canals in the later stages of distribution is connected with ease of mechanized cultivation.

Associated with irrigation, several wide forest-belts are being established, and smaller shelter beds encouraged over a wider area. The method of planting is shown in Inset 18–19. One of the purposes of the low-lying belts is to reduce surface

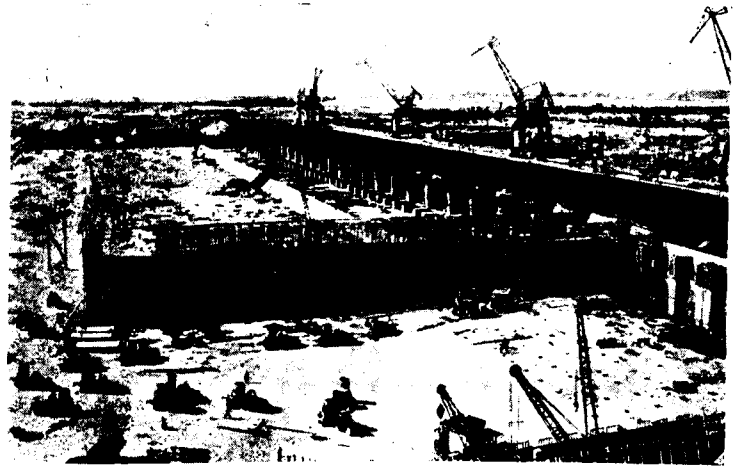
HYDROELECTRIC AND IRRIGATION DEVELOPMENT IN THE U.S.S.R.

1. Stalingrader-2, the most powerful suction-dredge yet completed. Designed to remove 10,000 cubic metres of soil per hour, it is stated to have reached 13,000 c.m./h. during test. It is electrically operated. (*This and the following eighteen pictures are reproduced by permission of* Soviet Weekly.)

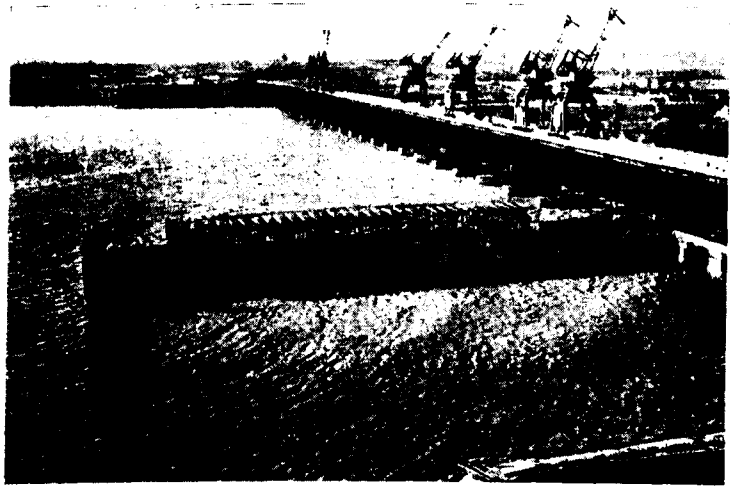
2. Tsimlyanskaya site (River Don): this and Inset 4 are reproduced from a single wide-angle picture. The natural bed of the river is seen at top left. Below is part of the excavated hollow leading to the spillway.



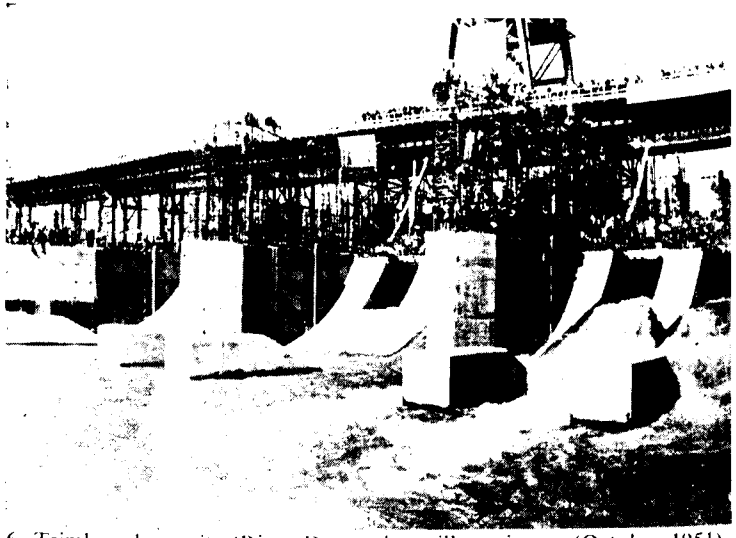
3. Tsimlyanskaya site (River Don): filling the spillway hollow. After river-level had been reached, and free access provided from river to spillway, blocking of the natural channel was completed: large quantities of concrete were dumped quickly in the final stages.



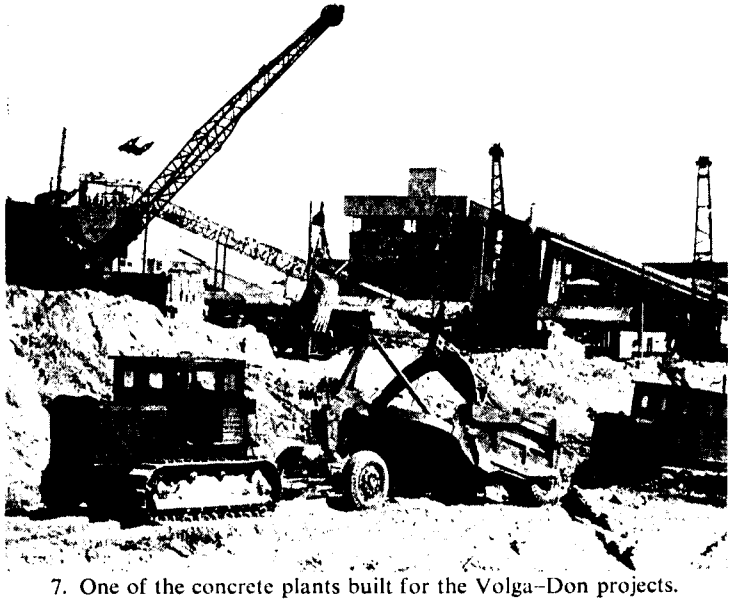
4. Tsimlyanskaya site (River Don): the spillway dam, still uncompleted in its upper courses, but ready for use at the natural river-level. Scale can be judged from the human figures seen in the foreground.



5. Tsimlyanskaya site (River Don): as Inset 4 *(above)*, but after filling to river level. The extent of backing-up which will be effected is indicated by the excess-height of the spillway dam. The appearance is of 'dead' water at this stage.



6. Tsimlyanskaya site (River Don): the spillway in use (October 1951).



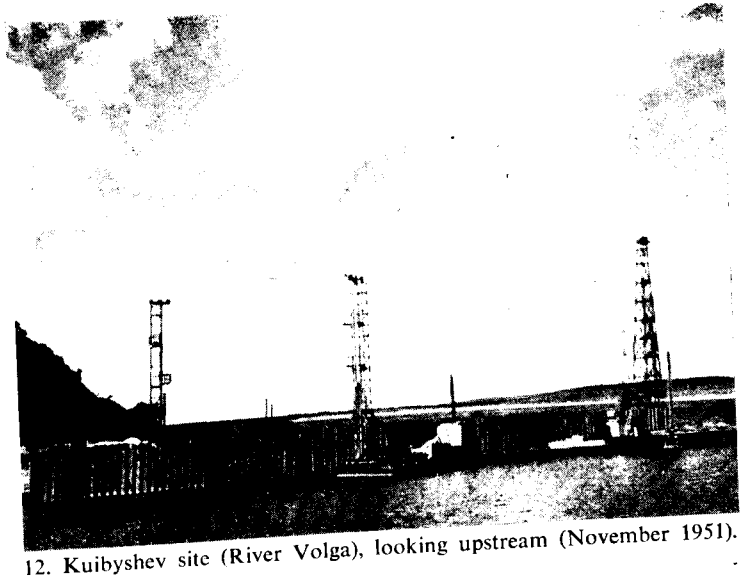
7. One of the concrete plants built for the Volga–Don projects.

8. Volga–Don navigation canal: dredging near the Volga outlet of the canal (June 1951).

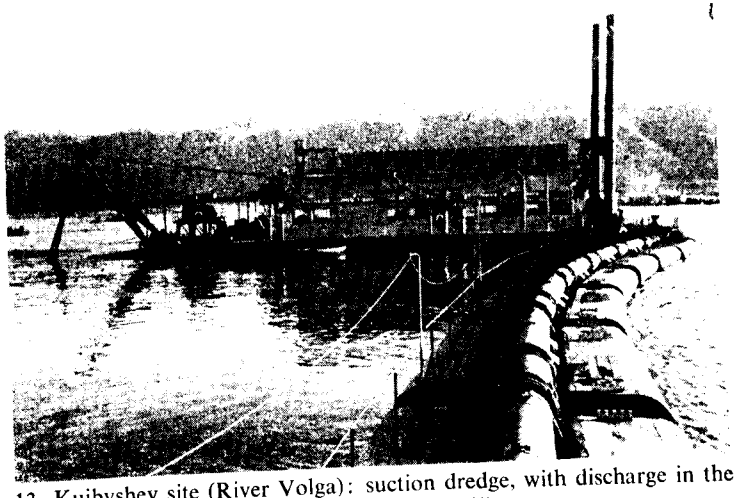
9. Volga–Don navigation canal: watershed section. Inspecting an underwater filter in the canal side (February 1952).

10. Volga–Don navigation canal: suction dredge at work (December 1951).

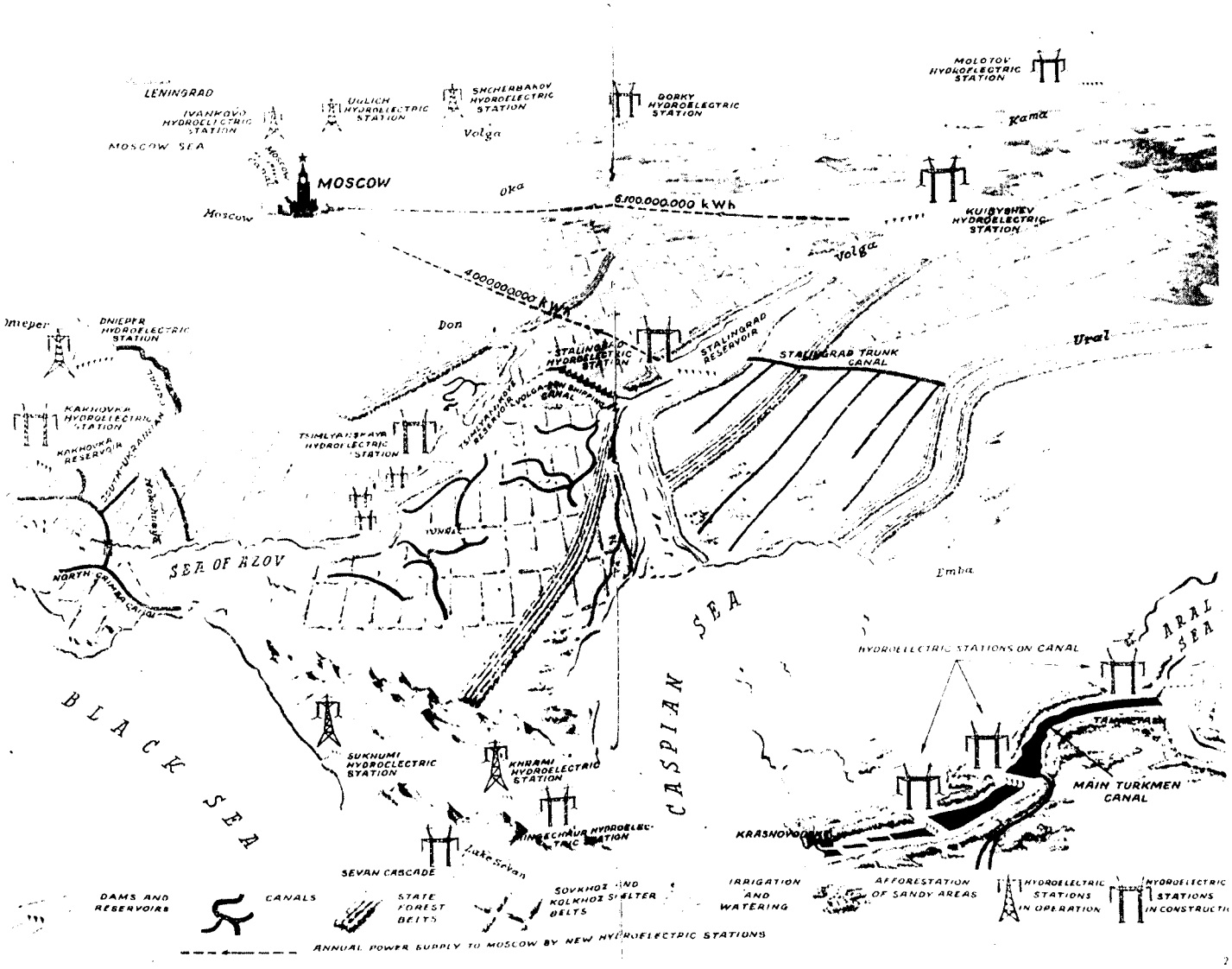
11. Kuibyshev site (River Volga): floating pile-drivers at work off the high (west) bank (August 1951). Work here is at an earlier stage.



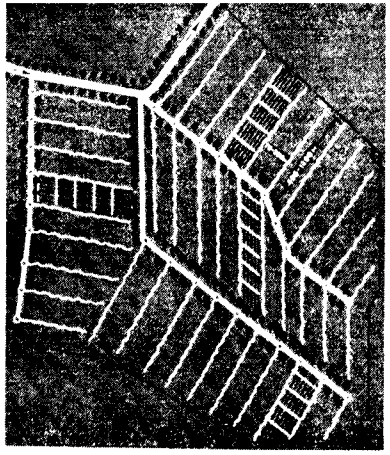
12. Kuibyshev site (River Volga), looking upstream (November 1951).



13. Kuibyshev site (River Volga): suction dredge, with discharge in the direction of the low (east) bank (December 1951).



14. Panoramic map showing the development schemes as a whole. As an indication of scale, the distance between the Gorky and Kuibyshev sites on the Volga is some 750 miles by river; that between Kuibyshev and Stalingrad about 500 miles. The distance, from east to west, covered by irrigation schemes is about 900 miles.



trunk canal

lateral canals

temporary (one-season) canals

feeder ditches

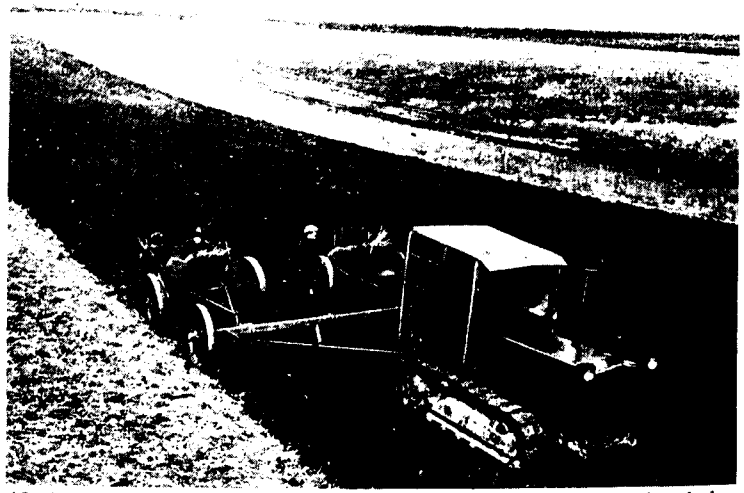
watering furrows

canalside plantations

15. Irrigation system using temporary canals. Only trunk and lateral canals are permanent. Within the individual irrigation unit, distribution is by temporary (one-season) canals, of 400–1,200 metres length, at intervals of 70–200 metres; and thence by feeder ditches and watering furrows. The system is adapted for use with mechanized cultivation.

16. Irrigation furrows: first watering of cotton field.

17. Tunnel on Don Trunk Canal. Fed from the Tsimlyanskaya reservoir, the canal will supply water for the irrigation of 1,850,000 acres, and will provide water for livestock over a wider area. It will be in four parallel tunnels, each 6 km. long, beneath the Don–Sal watershed.

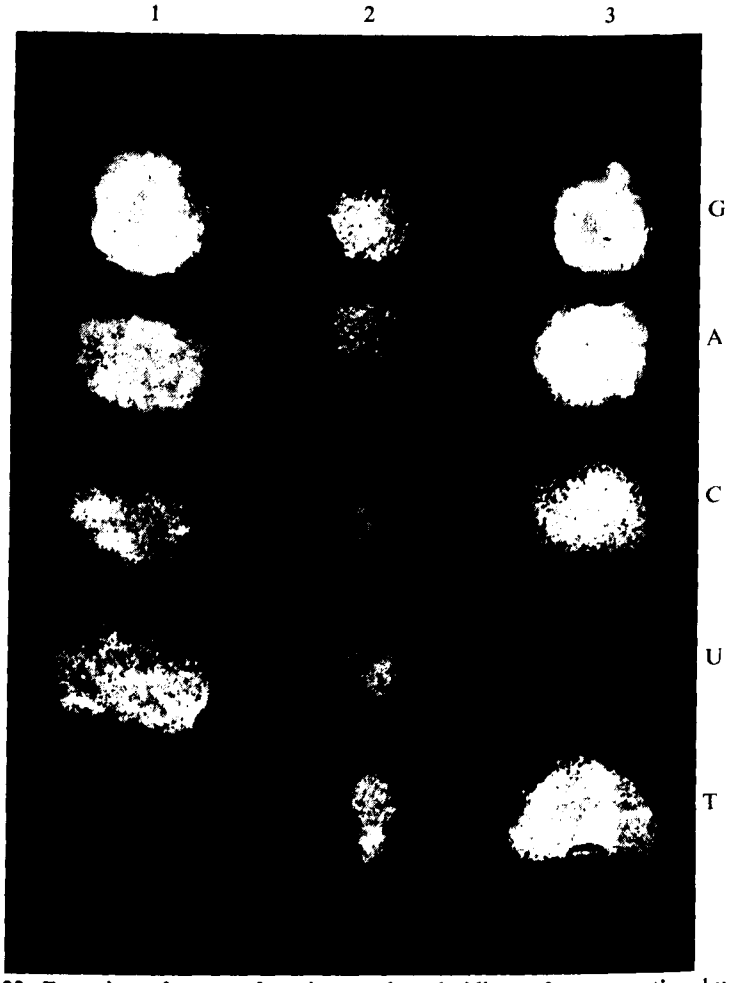


18. Use of mechanized tree-planting in the establishment of shelter-belts (May 1949): 7-unit machines, covering a 50-ft belt, are now in use. By breaking up the terrain, it is hoped to reduce surface evaporation by hot steppe winds.

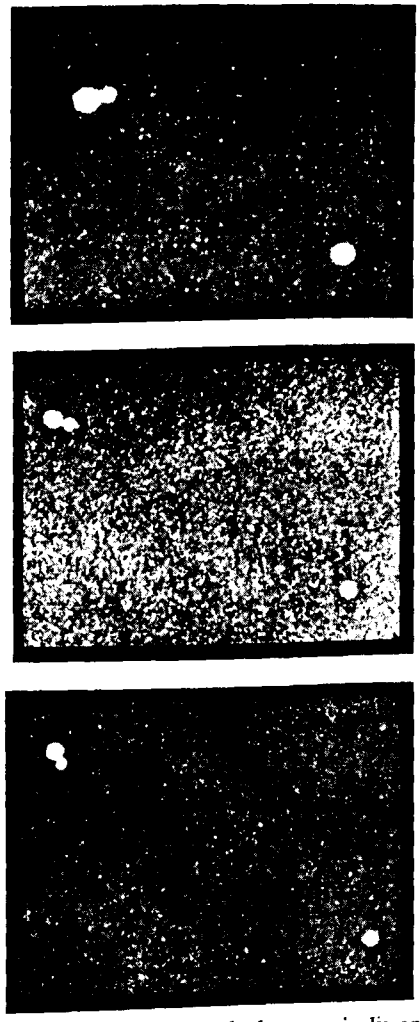
19. Experimental shelter-belts at the Dokucheyev Scientific Research Institute of Agriculture of the Central Black-Earth Zone. (*This and the preceding eighteen pictures reproduced by permission of* Soviet Weekly.)

20. Photomicrograph of nuclei isolated from calf thymus (× 1,500: unstained). (*This and the two following pictures are by F. Bradley.*)

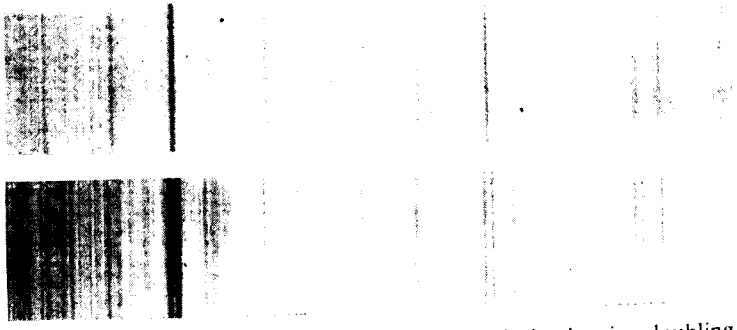
21. Fibrous deoxypentosenucleic acid (isolated from human spleen).



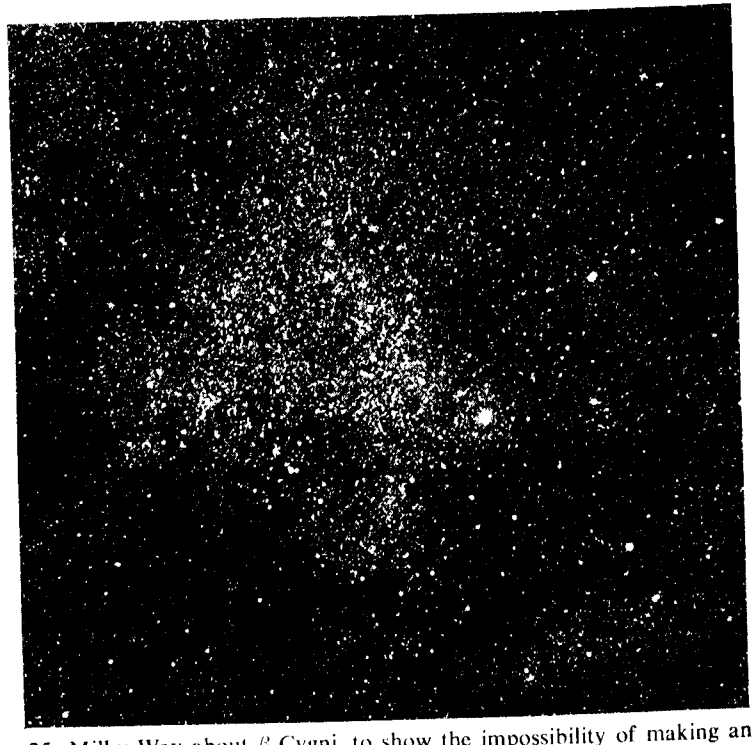
22. Detection of spots of purines and pyrimidines after separation by paper chromatography: photograph taken by ultra-violet radiation, with the sensitive paper underneath the paper strip. (1) Spots of the purines, guanine (G) and adenine (A), and the two pyrimidines, cytosine (C) and uracil (U). (2) Control strip showing positions of guanine (G), adenine (A), cytosine (C), uracil (U), and thymine (T), separated from a mixture containing 8 μg. of each. (3) Spots of the purines, guanine (G) and adenine (A), and the pyrimidines, cytosine (C) and thymine (T), from 0.5 mg. of DNA.



23. The binary Krüger 60, photographed successively on 21 July 1908 22 September 1915; 10 July 1920. *(This and the following two picture are reproduced by courtesy of Yerkes Observatory and the Universit of Chicago Press.)*



24. Spectroscopic binary: spectra of ζ Ursae Majoris, showing doubling of lines in the lower spectrum.



25. Milky Way about β Cygni, to show the impossibility of making an exhaustive study of visual binaries.

evaporation by hot, dry winds: they are, in fact 'wind breaks' used for a different object. A 15-year programme of state afforestation (as opposed to local planting by collective farms) is reported to be about one-third completed as regards planting. There has been speculation [*Weather*, **2**, 39 (1952)] as to 'what local climatic changes can be measured in years to come and whether they will, as suggested, extend to the climates of other countries bordering on the Soviet Union'.

On the hydroelectric side, the most important river is the Volga, on which there are existing stations at Ivankovo, Uglich, and Scherbakov (Inset 14). Below these, the scheme provides for the development of further stations near Gorky (formerly Nizhni Novgorod), Kuibyshev (formerly Samara), and Stalingrad. They are due for completion in 1955–56, and work on them is proceeding in parallel. The site illustrated is Kuibyshev (Inset 11–13). Dredging has also been begun at the Kakhovka site on the lower Dnieper scheduled for a year later. The most interesting point in the plan for this section is the proposal to carry an irrigation canal to the North Crimea across an arm of the Azov Sea.

Compared with most hydroelectric schemes, the Russian alluvial rivers present conditions of large flow, small gradient (averaging 2.16 inch per mile on the Volga at low water), and absence of rock foundations. The heads which can be obtained are therefore low relatively, and dams are long in proportion to height: for example, the spillway dam shown in Inset 3–6 is 0.5 km. long, but is flanked by earth dams with minimum revetting which total 13 km. length. The technical problems are generally similar to those encountered in the Punjab rivers, with the exception that floods, though serious, are less drastic. With increasing head, the requirements are to provide a broad enough base to support the main structure, and to reduce the effective gradient between the beginning and end of the dam-base to the extent necessary to ensure that interstitial flow beneath the dam is within the limit for stability. The effective gradient can be reduced by driving piles or other forms of baffle so that the path-length for the water is increased. In

relation to conditions, some of the dam-heights planned for the Volga appear formidable, notably at the Kuibyshev site. The reservoir-lake to be created by the Kuibyshev dam is planned to extend to above Kazan on the main river, a distance of some 370 miles. At the gradient already quoted, this implies a dam-height of about 80 ft. The generators are to be of 2,000,000 kW. rating, fractionally greater than that at the Grand Coulée dam on the Columbia River in the United States; this corresponds approximately with the utilization of the mean flow of the Volga, with the reservoir at its highest level, whereas under natural conditions between one-half and two-thirds of the annual flow is concentrated in the spring floods. As will be seen from the map, the amount of energy available for external use in an average year is estimated at 6,100 MkWh. (million kilowatt hours), representing an external load-factor of 35 per cent. The intended policy is to obtain as much energy from the Volga as possible, but in dry seasons to give priority to irrigation. The scheme as a whole is estimated to provide an extra 23,000 MkWh. of electrical energy per year, an increase of 40 per cent on the amount generated in Russia in 1940, and of about 25 per cent on 1950.

On the navigational side, the Volga–Don canal provides continuous access by water from the Baltic to the Caspian and, through existing canals, to Moscow, the Baltic and the White Sea. Completion of the Volga dams, later in the scheme, should relieve the difficult conditions, resulting from changes in the river-bed produced yearly by floods, as well as providing long sections of effectively lake-transport.

A. W. H.

STATISTICS AS AN AID TO LITERARY STUDIES

C. B. WILLIAMS

ABOUT ten years ago, in the course of general reading, I was struck by the way in which some writers tend to use rather long words, and others rather short ones; and it occurred to me that it should be possible to write a detective story in which the arch-villain, who only communicated with his victims in writing, was finally unmasked by the relative frequency with which he used shorter or longer words. I went so far as to count the frequency distribution of words of different lengths in works by several writers, including for example H. G. Wells, who used rather long words, and Lord Dunsany, who used rather short ones, but I came to the conclusion that the differences were too small to be of real value in determining authorship, so I laid the scheme on one side. A few years later I was interested to discover that a similar idea had occurred to the late Dr Udney Yule, a well-known statistician at Cambridge, but that he had counted, not the number of letters per word, but the number of words per sentence. This, with a very much greater range of variation – from one or two words up to a hundred or more – gave greater possibilities for comparison. And the differences were not only in the average length of the sentences, but also in the range of variation round the mean. Some writers habitually write rather long sentences, some rather shorter; some keep a more constant length, others are more variable. For example, in samples of 600 sentences which I took from the sociological writings of Shaw, Wells and Chesterton, Shaw ranged from 3 to 142 words, Wells from 2 to 91, and Chesterton from 5 to 91 but with only two sentences over 60 words (Fig. 1). The mean sentence lengths of Shaw and

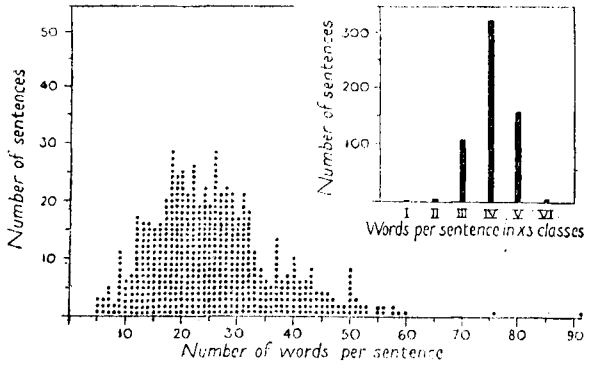


Fig. 1. Frequency distribution of sentences with different number of words in 600 sentences selected from *A Short History of England* by G. K. Chesterton: left on number basis; right in classes of × 3.

Chesterton were almost identical, but they could be distinguished because Shaw ranged further afield both in long and in short 'thoughts' (Fig. 2). Yule showed that the frequency distribution of sentences of different length was always 'skew': it had a much longer tail at the long end than at the short end, and so the average was much nearer the bottom than the top. My own small contribution to the problem at this stage was to show that if the sentences were grouped in lengths in geometrical proportion instead of arithmetical, the distribution became almost symmetrical, and, as a result, comparisons were much simplified.

Using his new technique, and with a more serious approach than mine had been, Yule tackled some problems of 'disputed authorship', and the following is an example of his method. In the early years of the fifteenth century there was written a manuscript entitled *De Imitatione Christi* (*Of the Imitation of Christ*). For many years there has been a controversy as to its authorship, various experts attributing it either to Thomas à Kempis (who lived 1379–1471) or to Jean Charlier de Gerson (1363–1429). Yule counted the lengths of about 1,000 sentences

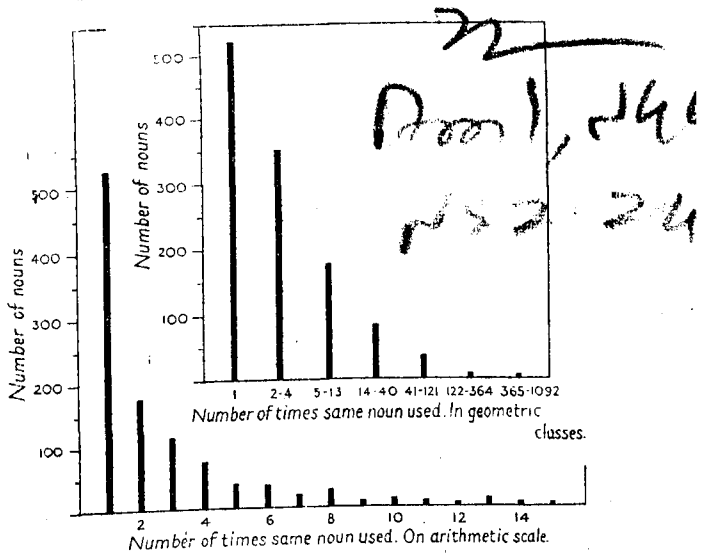
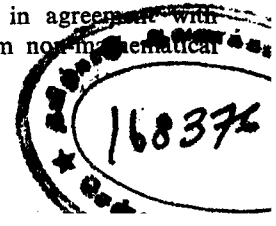


Fig. 2. Frequency distribution of nouns used different numbers of times in *De Imitatione Christi*: left, on an arithmetic scale (shown up to 15 only); right on a × 3 classification.

selected at random from *De Imitatione Christi* (in the Latin original), and then made similar studies from known works by each of the two candidates for authorship. He found that the mean length of sentence of the disputed MSS was 16.2 words, of works of Thomas à Kempis 17.9, but of Gerson 23.4. The variation of length, as measured by the range of the middle half of the sentences, was 9.2 words in the *Imitatio*, 11.8 in Thomas à Kempis, but 18.2 in Gerson's works. Other factors were also taken into consideration, and Yule concluded his study with the statement that 'These results are completely consonant with the view that Thomas à Kempis *was*, and Charlier de Gerson *was not*, the author of the *Imitatio*'. Although Yule's opinion was in agreement with the solution now generally accepted from non-mathematical



arguments, it was based on a new type of evidence and so was a definite contribution to the study of disputed authorship.

More recently W. C. Wake, in an article published in the *Hibbert Journal* for 1948, has applied this statistical method to the Pauline Epistles (using the Greek text), and finds distinct evidence of dual authorship. To summarize very briefly, he concludes that I Corinthians, II Corinthians, chapters 10–13, Galatians and Romans, have sentence lengths which are statistically indistinguishable, with a mean of just over 11 words. On the other hand, I Thessalonians, Colossians, Philippians and possibly II Thessalonians, from a second group with a mean of nearly 17 words and a much higher proportion of long sentences. The two groups almost certainly come from different authors and the first group is – for other reasons – believed to be authentic writings of St Paul. Certain characteristics of style and wording were found to differ in the two groups when they had been separated by statistical reasoning.

Meanwhile, Yule had gone off on another tack. He became interested in the diversity of vocabulary used by individual writers, and developed a method of measurement in which he counted the number of different nouns used in a sample of a large number of nouns in a particular piece of writing; and also how often each one was used. For example, from Bunyan's *Pilgrim's Progress* he took selections containing just over 4,000 nouns and found that they included just over 1,000 different ones. Of these, about 500 nouns were used only once, 179 were used twice, 87 three times and so on, and one – the word 'man' – was used 197 times. The frequency distribution was of the 'hollow curve' type, that is to say the number of nouns used steadily lessens in relation to the frequency of usage.

Yule was specially interested in what he called the 'Treasure Chest' of words in the mind of a writer. From the recesses of memory come up the words necessary for speech or writing; the well-educated man uses many different words by which he can express shades of meaning; the poorly-educated man uses comparatively few. The diversity of the mind is reflected in the diversity of the writing. To try to find the number of words in

a man's vocabulary Yule estimated (by a random sampling method), that there were about 56,000 different nouns in the *Shorter Oxford English Dictionary*, and of that number he himself was acquainted with about 20,000! Yule was, of course, a very widely educated elderly man. The number of adjectives in the *Shorter Oxford Dictionary* is barely half that of the nouns, but still amounts to about 27,000. It is interesting in this connexion to note that the man who has only one adjective is probably neither blasphemous nor particularly emphatic, but merely ill-read, ill-educated and rather boring – mathematically speaking 'undiversified'!

Yule applied his new technique to the problem of *De Imitatione Christi* (Fig. 3) and once more came down heavily in favour of Thomas à Kempis as the author.

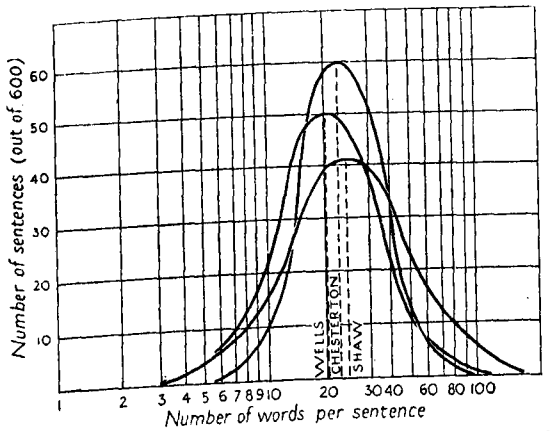


Fig. 3. Comparison of frequency distributions of sentence-lengths on logarithmic scale in samples from similar works by H. G. Wells, G. B. Shaw and G. K. Chesterton, showing differences in mean and in range.

It is not possible to go into details and difficulties; to cover the field, things must be simplified, but difficulties there are in profusion, and conclusions are never absolute. For example,

what is the definition of a sentence? If we define it as the words between two full-stops, who put in the full-stops? The author or the printer? Or even the printer's devil? One of the earliest of gremlins. Did the author correct the proofs? And, if so, did he in the process cut down (as I have done myself before now) the obviously long sentences, and so upset the original and natural order?

A still more fundamental difficulty is that before we can compare different authors, we must know what is the range of variation 'within an author'; that is to say, how he himself may vary according to the subject on which he is writing, his varying moods, and at different ages. Undoubtedly, as a man gets older his vocabulary will extend, and hence his writing will – up to a point – increase in richness and diversity. Recently this latter problem has been studied from a slightly different statistical angle by Yardi in India in connection with the chronology of Shakespeare's plays. He studied the number of lines with extra final syllables, and the number of split lines, in each of the plays. For those plays for which the date of writing was known he found a definite time relationship – for as Shakespeare grew older his vigour of mind and ideas overflowed more and more the bonds of metre. Yardi was then able by a process of interpolation to fit some of the plays of uncertain date in their most likely position in the sequence – and so produce new evidence of their chronology.

At a meeting in London about a year ago, at which problems of these types were discussed, Dr Ross, Professor of English Language in the University of Birmingham, expressed an opinion that he would sooner trust to the evidence obtained from verbal characteristics and forms of phraseology in determining authorship, than to statistical arguments. There is, of course, no doubt that in the hands of an expert like Professor Ross the unmeasurable subtleties of language can be summed up in a way that no mathematician can ever hope to achieve – but even then different authorities often come to different conclusions.

Where the record of authorship has been lost, or where there has been no effort to conceal or mislead, these subtle characteristics might be an excellent guide. But what happens when a writer deliberately writes a document anonymously? – or still more when – for good or evil motives – he plans to imitate the writing of another? Literary characteristics are not difficult to imitate, and so in cases of fraud may easily lead one astray. After all, we have recently seen an example of wholesale forgery of old masters in painting by a clever artist who knew what the experts would look for! But the thousand and one minute individual characteristics that are analysed statistically – the frequency distribution of different nouns; the mean length and variability of sentences – which each of us possess but of which we are mostly unconscious – these are too difficult to imitate, and so may be better guides to authorship than turns of the pen which a clever man could copy.

It has been said that Shakespeare was such a great playwright that no one knows what he really thought himself. Every character is so sincere in what he says that we know only the opinions of the puppets and not of their creator. I wonder how this portrayal of character and of individual mind would stand up to a statistical analysis; not of the things said, but of the way they are said. The poor writer gives his characters exaggerated verbal peculiarities, but the clever writer disdains such tricks, and hopes to make his creations live and have an individuality without them. How well does he succeed? May I suggest that someone should make a study of sentence length and word frequency distribution in, say, the Hell scene in *Man and Superman*, keeping the speeches of each character separate. Then he could see if Shaw was really so great an artist that each of his characters spoke from its own mind and memory of words and thought, or whether they merely drew their vocabulary from the mind of their creator. In other words was Shaw 'Superman or Man'?

Before concluding, I must emphasize that it is not suggested for one moment that the statistical method should replace any other approach to literary problems, but that it should be used in addition to them. It is a new weapon, just as a microscope

was once a new weapon to the biologist, and those who despise or neglect it will have only themselves to blame if their work suffers. I began by saying that it should be possible to write a detective story on this thesis. But what, after all, is the research that I have been describing but a glorified detective story? There is the search for clues, the separation of red from other-coloured herrings, analysis and deduction, all the fun of discovery without the sordidness of crime. Even the end of a piece of research is much like that of a detective story; the last chapter is merely an inducement to read the next volume! What, I wonder, will be in the next volume of this problem I have just outlined – and who will write it?

REFERENCES

Ross, A. S. C. *J. Roy. Stat. Soc.* B **12**, 19 (1950).
Wake, W. C. *Hibbert Journal* **47**, 50 (1948).
Williams, C B. *Biometrika* **31**, 356 (1940).
Yardi, M. R. *Sankhya Ind. Journ. Stat.* **7**, 263 (1945).
Yule, G. Udny. *Biometrika* **30**, 363 (1939).
Yule, G. Udny. *The Statistical Study of Literary Vocabulary*. London, 1944.

TAKING A SAMPLE

J. A. NELDER

IN the last half-century, and particularly in the last twenty-five years, the theory of the design and analysis of scientific experiments has made great strides, and some knowledge of it is rapidly becoming indispensable to working scientists. Because of the usual time-lag attending the application of new discoveries, the concepts employed in *statistics*, as this subject is usually called, are unfamiliar to most laymen and to many scientists. The aim of this article is to present certain ideas that statisticians have developed, by considering a concrete problem, namely that of drawing a sample from a large collection of objects, otherwise from a *population*, in order to gain information about the whole by examinaiton of a part only. The objects in question may be such things as plants in a field, leaves on a tree, bacteria in a liquid, or manufactured articles on a production line.

Suppose that the population in question is that of potato plants in a field. It is obvious that we cannot make exact statements about, say, the mean yield of all the plants in the field if we look at only a portion of them. Our inferences about the whole from the part will be more or less uncertain; this is fundamental to statistics of the kind we are talking about. [The term is also used in connexion with the compilation of complete sets of data, such as trade returns, into which this uncertainty does not enter: we shall not consider this type of statistics.] Statistics aims at supplying a sound basis for uncertain statements. It may be asked why, in the 350 years or so of 'modern' science, statistics should have reached prominence only since 1900. The answer is not far to seek; in a branch of science like early physics, by using a careful technique, quantities could often be measured with only a comparatively small

error, or at least with an error which could be made small in relation to requirements at the time. A measure of the degree of uncertainty was not vital. But in the biological sciences in particular, experiments are beset by uncontrollable influences having effects comparable in size with the experimental factors. The effects of soil fertility and of pests in field experiments are examples of such influences. Here it is obviously necessary to have some technique for distinguishing, wherever possible, between real differences, *i.e.* those due to the effect of the factor under trial, and 'error' or spurious effects due to the haphazard influences of a multitude of factors outside the control of the experimenter. Techniques for dealing with this situation come under the heading of the design of experiments and will not be considered here. The same basic theory, however, covers both the design of experiments and sampling methods.

What is the effect of these uncontrolled influences on a population, such as our potato plants? Clearly it is to introduce variation. Time of planting, soil differences, genetical differences and so on will all tend to introduce variation between one plant and another. To demonstrate the characteristics of such variable material, the writer has taken a sample of 90 potatoes from a sack and weighed each one separately to the nearest ½ oz. Fig. 1, called a histogram, shows graphically the number of individuals in each category, grouped in units of ½ oz. The actual number of individuals in each group is shown as a figure in the corresponding rectangle.

The histogram gives the frequency with which the weight, a

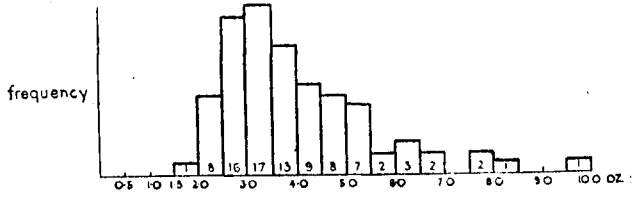


Fig. 1. Histogram of weights of 90 potatoes.

variate, takes values in a certain range in the sample. The assemblage of frequencies forms a *frequency distribution*. Out of 90 potatoes, 45 have weights lying between 2.5 and 4 oz. Hence, if we take one at random from the sample, the chance is 46 out of 90 (just over ½), that its weight will lie between 2.5 and 4 oz. Similarly the chance is 61 out of 90 that its weight will lie between 2 and 5.5 oz.

Some frequency-distributions found in practice are symmetrical about the largest or modal frequency. Others, like the one above, are skew and have a longer 'tail' on one side than the other.

Corresponding with the empirical frequency distribution, the statistician uses a theoretical construct known as a *probability distribution*. Roughly speaking, a probability distribution is a smoothed-out frequency distribution in which the variate (corresponding with the weight in our frequency distribution) takes values with a probability following some mathematical law. Fig. 2 shows the most important of these, the normal or Gaussian distribution.

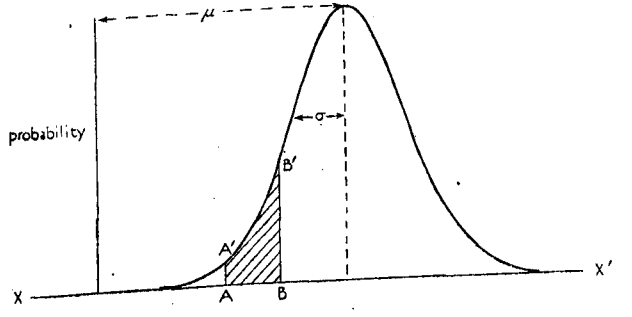


Fig. 2. The normal distribution.

The line XX' is the axis of a continuous quantity, or variate, x say, which can take any positive or negative value. A partial area, such as $AA'B'B$ (shaded in the figure) represents the probability that x lies between A and B. The total area under

the curve, representing the probability that x lies somewhere on the line, is unity, the conventional number used to denote certainty. The normal curve is completely determined by two quantities or *parameters*; the *mean*, usually denoted by the Greek letter μ, which is the value of x corresponding to the top of the hump (M in the figure), and the *standard deviation*, usually denoted by the Greek letter σ, which is the distance between the mean and a point of maximum slope on the curve. The shape of the normal curve is given by the equation:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

in which $e = 2.718\ldots$ is the base of natural logarithms. From this law it can be shown that 95 per cent of the area of the curve lies within 1.96σ of the mean and 99 per cent within 2.58σ of the mean. The quantity σ^2 is called the *variance* of the distribution and is the average value of the square of the deviations of x from its central value or mean.

After this necessarily brief outline of frequency and probability distributions, we now return to the problem of drawing a sample. In order to be able to generalize from our sample to the whole population we should like the sample to be representative. That is, if we draw a sample of, say, 50 plants from 10,000 we should like the variation in the sample to mirror on a smaller scale, so far as possible, the variation in the whole population. It might be thought that we could pick a typical set of plants from the field simply by eye; unfortunately, it has been shown beyond all doubt that human judgement in this respect is very fallible and liable to *bias*. By this we mean that experimenters often have a tendency to select plants that are on the average too large or too small in their yield (or whatever other character we are sampling for). We can eliminate this bias by taking a *random sample*. To do this, we use an extension of coin-tossing and make our selection in such a way that each plant in the field has an equal probability of being selected. The most convenient way of doing this is to number

off the rows, and then the plants in a row, and to use a table of random numbers to make the actual selection. Several sets of random numbers have been published; they have the property that each digit occurs about an equal number of times in a long series, but locally the series is quite haphazard.

Let us suppose that, using a random sampling technique, we draw a sample of 50 from 10,000 plants in a field, and that we wish to estimate the total yield from the yield of plants in the sample. This we can do quite simply by adding up the yields of the plants in the sample and multiplying by 200 since we have sampled 1/200th of the population.

Can we now get an idea of the possible error of our estimate? It is the second great advantage of the random sampling technique (the first being the lack of systematic bias) that the answer to this question is yes. Since each plant in the field has an equal chance of being chosen, there are

$$\frac{10{,}000 \times 9{,}999 \times \ldots 9{,}951}{1 \times 2 \times 3 \times \ldots \times 50}$$

possible combinations of plants in a sample of 50, each equally probable. Each such possible combination would give an estimate of the total yield. The frequency distribution formed by all these possible estimates is called the *sampling distribution* of the estimate. The sampling distribution can be shown to be very well approximated to by the normal curve when the estimate is the sum of as many as 50 yields even if the original distribution of single yields was not exactly normal. The mean value of this normal curve is the quantity we are trying to estimate, namely the true total yield. The estimated value we get from the sample will be different from this true value but it will lie within various distances of it with various probabilities, depending on the standard deviation of the curve. If the plants are very variable in their yields, this standard deviation will be large and our estimate may quite easily be rather far away from the true value. With a uniform 'stand', the probability of a large deviation from the true value would be much reduced. However, we do not know the standard

deviation of the sampling distribution, but it is characteristic of the random sampling technique that, although we have only one value of the sampling distribution of our estimate, we can estimate the variance (and so the standard deviation) of this distribution from the variation *within* the sample.

Let the yields of the plants in the sample be $x_1, x_2, \ldots x_{50}$ and let our estimate of the total yield be M so that $M=(x_1+x_2+\ldots+x_{50})\times 200$. The variance of a single yield is estimated by the quantity $S\,(x-\bar{x})^2/49$, in which $\bar{x}$ denotes the mean of the x's, S denotes summation over all the x's, and 49 (the divisor) is one less than the number of yields in the sample. The variance of the sum of k independent variates, each with the same variance, is k times the variance of any one of them. Thus the estimated variance, of $x_1+x_2+\ \ldots\ +x_{50}$ is $S(x-\bar{x})^2.50/49$. Finally, if we multiply a variate by k, its variance is multiplied by k^2, so that the variance of M is estimated by $(200)^2.S(x-\bar{x})^2.50/49=s_M^2$ say. Knowing s_M, we can now make statements of the kind – 'If the true total yield is Y, the odds are 19 in 20 that our sample estimate will not deviate from Y by more than $2.01s_M$. It will be seen that these limits are slightly wider than those given for the normal distribution. This is because we are comparing the variate with an estimate of its standard deviation and not with the (unknown) true standard deviation. The slightly greater limits of uncertainty may be regarded as a consequence of the inexactness of our information about σ_M. The distribution of the ratio of a normal deviate to an estimate of its standard deviation is known as 'Student's' t-distribution, after the discoverer 'Student,' the pseudonym of W. Gosset. It is from this distribution that the limits $2.01s_M$, mentioned above, have been calculated.

The type of statement above does not cover a much more important situation: namely, given the sample estimate, what can we say about the probable limits of the unknown true value? This is the inverse problem to that solved above and, in its various forms, it has given rise to considerable controversy among statisticians. We have only space to consider

one approach (first made by J. Neyman) which provides *confidence limits*, as they are called, for the unknown value, given the sample estimate. In our case, we have a variate, namely a sample estimate of the total yield, M and an estimate s_M of the variance of its sampling distribution. Neyman showed that, given a fraction p, the statement 'the real value, Y, of M lies between $M-t_ps$ and $M+t_ps$' will be true in a proportion p of cases in the long run; t_p is a quantity depending only on p and the number of readings used to estimate s_M, and not on the unknown parameters of the sampling distribution. For $p=0.95$ and a sample of 50, $t=2.01$ and is the same quantity as occurred above in the other problem. For $p=0.99$ and a sample of 50, $t=2.68$. If, for example, our estimate M came out to be 10 tons and s_M to be 0.5 tons, then using the 95 per cent level, we should say 'the real value of the total yield lies between $10-(2.01\times 0.5)=9$ tons and $10+(2.01\times 0.5)=11$ tons' knowing that if we took a series of samples and made such a statement about each estimate we obtained, we should be right in 95 per cent of the cases in the long run in so doing. If we wished to be right 99 times out of 100 the confidence limits for the real value would, of course, be wider.

In order to bring out the salient points of the random sampling technique we have used the simplest possible method, and this has led to some lack of realism in the example chosen. For example, on common-sense grounds we should expect it to be better to divide the field into portions and to sample each portion separately. In this way, we should be sure that any large patches differing from the rest would be adequately represented. An actual sampling technique would be likely to deviate from the elementary type considered above by various such modifications designed to improve accuracy (*i.e.* to reduce sampling variances) while retaining the properties of lack of bias and availability of valid estimates of error which are characteristic of the simple random sample.

There has only been space in this article to touch on one aspect of a rapidly growing subject. Reaching as it does to the fundamental problems in science of induction and

generalization, statistics and the concepts associated with it are bound to make their influence felt wherever scientists work. At present, the statistician finds great difficulty in explaining to a layman the scope of his job. This is chiefly because of the barrier caused by the esotericism of the language of statistics. It is hoped that this article may help a little in breaking down this barrier.

RESEARCH REPORT

A. W. HASLETT

ICE, RAIN, AND ELECTRIC CHARGE

THE freezing of water has been studied in a number of laboratories since the war in relation to rain-formation. One line of research has been on the extent of supercooling that occurs in water-droplets, and on the number of ice-forming nuclei in air that become active as such, as the temperature is progressively lowered. A second, and more recent, line of research has been on the electrical effects which result from the freezing of water onto ice.

The interest of the first line of research is in relation to one of the basic processes leading to rainfall. For rain to fall, it is necessary that the small initial droplets, formed by cooling in the rising air of a cloud, should coalesce sufficiently to fall through the rising air-current by which the cloud is maintained. Because the vapour-pressure of water is greater than that of ice at the same temperature, an ice-particle, once formed, tends to grow at the expense of neighbouring water droplets. The formation of ice-particles near the top of the cloud is therefore one process, in temperate conditions probably the most important one, by which rain is caused to fall. The effect of laboratory experiments on ice-forming nuclei is to suggest that water-droplets freeze more readily at cloud-top temperatures (commonly at about -15° C.) than would be expected from the number of ice-forming nuclei then active. There may be a possible explanation in a chain-reaction which would enable a very few ice-crystals, once formed, to multiply rapidly. The supposed mechanism consists in the throwing off of splinters from the initial ice-particles, either when water-vapour condenses onto them or when water freezes to their surface. If that is a correct explanation, it would follow that the nuclei which are active at cloud-top would normally be

few in number and difficult to identify; and there would be only a small chance that further laboratory experiments in this direction would prove helpful to the meteorologist.

The electrical effects which accompany the freezing of water onto ice are of interest in relation to the charge carried by raindrops. Measurements of this charge have been made in a number of countries from about 1908 onwards, but have been mostly of the net charge carried by numbers of raindrops. From such observations, it can be said that most rain is positively charged on balance; snow and drizzle being predominantly negative, and thunder-rain more variable as to sign of charge. There have been relatively few measurements of the charge carried by individual raindrops. The most recent, and the most numerous, have been by Dr Ross Gunn of the U.S. Weather Bureau. He has used an apparatus in which the charge of an individual drop is measured by the current it induces in passing through a closed conducting loop: there are, in fact, two such loops, separated by a known distance, the time-interval between the pulses from the two loops giving also a measure of the speed of the raindrop, when conditions are otherwise suitable. The masses of individual drops are estimated from the size of the spots that they make on the blue-print paper used for recording. Such measurements have been made from an aircraft during flights through a cold front and thunderstorms, and also at ground-level during thunder-rain. Dr Gunn's general conclusion is that positively and negatively charged drops are present simultaneously in most parts of both cloud and rain, and in thunder-rain at ground level he found that only about two consecutive droplets on the average carried an electric charge of the same sign.

Observations of this kind are probably at a rather early stage for generalization, and in the case of those made from aircraft there may be possible difficulties in interpretation. Dr Gunn's view at this stage is that the same basic process of charging operates in most rain, and that the characteristic property of a thunderstorm is the greater separation of charge which takes place owing to the more rapid vertical movement

of air in the cloud. There is thus an extended interest for meteorologists in laboratory processes which could lead to a local separation of charge, as between one raindrop and another.

Experiments in which supercooled drops of water were allowed to fall onto an insulated cold metal plate were begun in 1946 by E. J. Workman and S. E. Reynolds of the New Mexico School of Mines, and resumed later with the drops of water replaced by drops of dilute solutions of various salts. More lately, E. W. B. Gill and G. F. Alfrey of the Clarendon Laboratory, Oxford, have returned to the simpler case of water freezing on ice, and have given a preliminary account of their conclusions, with most of the detailed evidence left for later publication. Their basic experiments were ones in which a cylindrical block of copper was first cooled in liquid air and then partially submerged in distilled water in a beaker: the base of the beaker was insulated, and the difference in potential between water and copper-block was measured with an electrometer. The greatest potential difference was about 50 volts, the ice (frozen onto the copper) being positively charged with respect to the water. The potential difference was greatest when freezing was most rapid, and was reduced as the freezing-rate was reduced. If the ice began to melt again, there was a reversal of potential at about the same time, but to the extent of only a few volts. The interpretation (based on further experiments not yet published) is that the contact potential between water and ice is relatively small, probably a fraction of a volt, and that the higher observed potentials are due to the freezing into the ice of successive charged layers, the corresponding negative charge remaining in the water. The smaller potential difference found at lower freezing-rates is explained by the leakage of charge through the ice; and the reversed potential by the release on melting of part of the positive charge which had been frozen into the ice, part of the initial negative charge of the water having in the meantime leaked away.

On this basis, it might be expected that the impact of water

droplets on ice particles would lead to an accumulation of positive charge on the ice-particles in the first instance, with the part of the water-droplets which rebounded left with a negative charge. The ice-particles, falling through the cloud, might lose part of their charge by collisions with water-droplets of higher temperature, so that partial melting was produced. However, ice-particles which had grown in the first place by repeated freezing could in no case experience a reversal of charge; and their eventual melting to form raindrops would account for the preponderantly positive charge associated with most rain. The occurrence of negatively charged drops could be explained either by the reversal of the sign of the effect by impurities, on which there is preliminary evidence from the New Mexico experiments, or possibly by some further process of which the negatively charged (splashed off) droplets provide the starting point.

Dobson, G. M. B. *Proc. Phys. Soc.* 'B', **63**, 252 (1950).
Gunn, R. *J. Geophys. R.*, **54**, 57 (1949).
Weather, **6**, 27 (1951).
Workman, E. J., and Reynolds, S. E. *Phys. Rev.*, **78**, 254 (1950).
Gill, E. W. B., and Alfrey, G. F. *Nature*, **169**, 203 (1952).

SOIL STRUCTURE

The recent announcement of a synthetic soil-conditioner, 'krilium', of which the effect in laboratory experiments is to stabilize the structure of soil-particles with which it has been mixed, is of considerable interest to soil science, apart from the practical applications intended for it. The origin of the research leading to 'krilium' is apparently an investigation of the possible agricultural uses of alginates (from seaweed) which was undertaken at the Rothamsted Experimental Station during the war by the Unit of Soil Metabolism directed by Dr J. H. Quastel for the Agricultural Research Council. With D. M. Webley, who had developed a microbiological method for measuring the aeration of a laboratory-constituted standard soil, Dr Quastel studied the effects of various additions to

this soil, including sodium alginate, cellulose acetate and related substances, wheat straw, and water-soluble extracts obtained from fertile soils and from peat. The implication from these experiments was that soil structure was improved in respect of crumb-stability by the addition of a variety of polysaccharides, which having a chain-like or net-like structure might be supposed to act by causing clay particles to adhere more firmly together. However, the experiments were unsuccessful in their immediate object of finding an economic use for alginates in farming practice, and they attracted little attention when published in 1947. There are indications from other research that bacterial polysaccharides may contribute significantly to soil-structure under natural conditions.

The Unit of Soil Metabolism was transferred to Cardiff in 1945, and two years later Dr Quastel was appointed to a chair in biochemistry at McGill University, Montreal. He took his ideas with him: notably that if synthetic substances could be made which showed similar effects, were active in small quantities, would remain unchanged in the soil during reasonable periods, and were non-toxic to plants and animals, then this might be the solution to the problem.

'Krilium', developed by the Monsanto Chemical Company, is described as a water-soluble resin, which in solution is a polyanion with 100 or more negative charge per 'ion'. No specific composition is published, but a polymer of acrylonitrile is mentioned as one of a number of substances which have been prepared experimentally and show the wanted activity on hydrolysis. Laboratory experiments have shown that soil-crumbs formed by working a clay soil treated with 0.1 per cent 'krilium' retain their structure when covered with water unless mechanically crushed, and that capillary action, water-holding capacity, and soil-aeration are increased by treatment. It seems clear, therefore, that clay particles are held together by 'krilium'; and that the effect in laboratory experiments is greater, weight for weight, than that produced by natural polysaccharides.

The mechanism of holding-together cannot be simply related

to existing knowledge of the distribution of electric charge on soil particles. Clay, repeatedly washed with acid ammonium oxalate in sunlight, loses about 7 per cent by weight of iron and aluminium, and is left as a nearly white residue which is not only negatively charged (as is 'krilium'), but carries a greater negative charge than does untreated clay. The first result shows that clay particles, as such, could not be held together by a negative network such as 'krilium' could provide. The second suggests that natural, untreated clay particles contain components which are positively charged (so accounting for the smaller negative charge associated with untreated clay), and that these components are hydrated oxides of iron and aluminium. It is supposed, therefore, that there is an equilibrium between positive ions in solution in soil and positive ions adsorbed to clay particles. The latter, if more than 1-valent, could link together the negative charges of elementary clay particles and the presumably, more extended negative network provided by 'krilium'. The study of electric charges on soil particles is one of considerable difficulty, and there is a further difficulty in relating the effects of natural polysaccharides on soil-structure to these basic experiments. Irrespective of its practical use (in which scale of use and cost must be directly related), there will be a value to research in the advent of a stable, synthetic equivalent of the natural polysaccharides, of approximately known structure.

Tests on an agricultural and engineering scale are still at an early stage. Suggested uses included the improvement of soil-structure in market-gardening and in private gardens, the control of erosion pending the establishment of protective vegetation, and in road-construction and similar work on clay soils.

Quastel, J. H., and Webley, D. M. *J. Agricultural Sci.*, **37**, 257 (1947).

'Technical Background Information on Krilium'. Monsanto Chemical Company (1951).

Schofield, R. K. 'The Electric Charge on Soil Particles'. *Rothamsted Experimental Station, Harpenden: Report for 1947*, pp. 95–100 (1948).

SELENIUM IN PLANTS

The results of a co-operative survey of selenium-containing plants native to South Dakota were presented by the United States Bureau of Plant Industry and Animal Husbandry at the recent annual meeting of the American Association for the Advancement of Science. Selenium is believed to be the only element which is toxic to animals, occurs in soils only in minute quantities, but is taken up by healthy plants in amounts large enough to make them poisonous. Certain plants, known as 'selenium indicators', are found always to contain selenium when grown on soils containing the element, and appear to require selenium for their growth. This group includes 24 species and varieties of *Astragoclus* (vetch), and all species of *Xylorrhiza* (wood aster), *Conopsis* (goldenweed), and *Stanleya* (Prince's plume). Plants in this group release volatile substances, characterized by the strong and unpleasant smell of organic selenium compounds. Preliminary experiments suggest that selenium, in cereals, is associated with protein; and that it is probably present in selenium-containing amino-acids analogous to the sulphur-containing amino-acids, cystine and methionine. Radioactive selenium is being used in 'tracer' experiments which it is hoped will give more definite evidence on the selenium-metabolism of these plants.

Bio-Sciences Newsletter, Vol. II, No. 2, p. 6 (1952).

CORRESPONDENCE

SCIENCE AND CULTURE

DR A. E. Bell wrote in a letter in *Science News* 19 (p. 138): 'Nothing short of a full study of man as an organism and his place in nature can equip us to take control of the forces of the modern world, and all that the religious teachings of the past seem to have inculcated needs to be superseded in these respects. These are controversial problems, and the writer would reject all suggestions that science can do the duty of religion or even provide any true basis for ethics.'

This is surely a contradiction in terms. Dr Bell appears to be preaching the religion which has recently been given the name of 'scientism', the belief of 'scientians'. (Lunn, *The Revolt against Reason*, Eyre & Spottiswoode, 1950.) What does Dr Bell propose will replace the religious teachings of the past once these have been superseded, and how does he propose to replace them if he 'would reject all suggestions that science can do the duty of religion or even provide a true basis for ethics'?

Whether we call the system of principles on which we base our conduct (including the use we make of modern scientific discoveries) philosophy or religion does not matter very much, for any system of ideas on which we act is a 'religion'. It does matter, however, that these principles be true. The question is how are we to arrive at such truth.

I am suggesting that the tool we must use is the reason, at least to lay the foundations of our systems of beliefs, i.e. we must without prejudice submit *all* available evidence (not simply that supplied by the natural sciences) to its intellectual scrutiny. Dr Bell, for instance, starts with the 'passion and prejudice' that only these disciplines – i.e. the natural sciences – can tell us the truth; or perhaps it would be fairer to say that

he starts with the prejudice that all religious teaching should be superseded. *Some* religious prejudices should certainly be superseded, just as some scientian prejudices need superseding when applied without discrimination to problems essentially philosophic.

Dr Bell also says (p. 138) that traditional teachings have no answer for problems arising all around us. Some traditional teachings do claim to have an answer. Would Dr Bell care to state those teachings which he considers supply no answer, and the grounds on which he considers them now worthless?

Dr Bell states on the same page that the strife of the modern world is the result chiefly of the survival of interests and institutions which live on the passions and prejudices of mankind. This statement contributes nothing to explaining why mankind has these passions and prejudices. And surely Dr Bell is open to the charge that his own beliefs are their product.

The review of Planck's book which follows Dr Bell's letter (p. 139) is pertinent to these questions. It needs pointing out, however, that if, as Planck apparently believed, religion operates predominantly with sentiment it is as valueless from a philosophical and practical point of view (e.g. in the field of ethics) as Dr Bell's scientism. Philosophical/religious belief must in the first place be based on rational investigation; reason is the common ground on which men can meet intellectually, and all religions, be they scientistic or deistic, must come before its bar initially. It will be clear, I hope, that I have no quarrel with Dr Bell when he proposes that as complete a natural scientific analysis of man and his nature as possible will contribute to his understanding of himself. But natural science is not the only true road to knowledge of reality. There is a body of modern philosophers who are of the opinion that reason is able to give us a grasp of certain aspects of this reality on which action can be safely based, but they insist that the field of investigation must be wider than that normally understood by natural scientists who philosophize today.

Dr Bell suggests that imagination, principally through science (he means natural science presumably – there are

others) can bring about an eventual rapprochement between cultural groups. I suggest that reason, with the help of natural science, will prove a more effective combination.

A. T. MACQUEEN

St Andrews University, Dundee

PICTURES IN THE MIND

DR J. A. V. Butler appears to give a very misleading view about the conduction of the heart impulse on p. 28 of his article 'Pictures in the Mind' in *Science News* 22. His description implies that the sequential stages of the heart cycle are dependent upon afferent nervous impulses originating in the heart at one stage of the cycle initiating, through a reflex arc, the following stage. While the heart rate, and perhaps the strength of the beat, can be modified through the action of reflexes originating in the heart itself and in the neighbouring blood vessels, such reflexes are not engaged in the actual propagation of the beat. The apparent similarity of some servo-mechanisms to biological activities can be pushed too far. It would appear that Dr Butler has never observed the beat of an isolated perfused heart or of the auto-perfused heart in a heart-lung preparation. In both preparations, the heart is separated from the nervous system and, therefore, no mechanism such as he postulates can operate.

In a following paragraph, the author refers to reflex actions produced in response to 'sensations'. Presumably he means 'stimulations', and one can recommend Sir Charles Sherrington's definition of reflex action to him.

A. HEMINGWAY

Department of Physiology, The University of Leeds

Dr J. A. V. Butler writes:

I am indebted to Professor Hemingway for his correction of the account of the mechanism of the heart-beat which I used as an illustration of an automatic type of control of muscular actions. It was, however, an incidental illustration and does not affect the main argument.

BOOKS RECEIVED

SCIENCE AND HUMANISM by Erwin Schrödinger (Cambridge University Press, 1951), pp. 68, 8s 6d.

In four public lectures, delivered in 1950 under the auspices of the Dublin Institute for Advanced Studies, Professor Schrödinger discussed the humanistic value of science (a part-answer, it is suggested, to Plotinus's question, 'who are we?'); the present state of physics, with 'discontinuity' viewed as a return, in part, to earlier ideas; and the relationship, if any, between physical 'indeterminacy' and 'the free-will problem' ('If there is such a problem, it is not furthered a whit by the latest development in physics'). Now published in book form, the lectures show the qualities of charm, breadth, and incisiveness; and, if some readers may wish to stop and argue in places, what better tribute?

ELECTRICAL PHENOMENA AT INTERFACES edited by J. A. V. Butler (Methuen, 1951), pp. 309, 32s 6d.

The descriptive sub-title 'in chemistry, physics and biology' gives the key to the conception and contents of this book. To biologists it presents physical concepts and illustrations; to physicists and physical chemists 'the more complicated systems which occur in nature'. A foundation on the physical side is provided by four chapters revised from Dr Butler's *Electrocapillarity* (1940). On the biological side, W. F. Floyd writes clearly and critically on membrane potentials, and nerve and muscle in particular. Other contributions are on electrokinetic effects, introduced by J. M. Creeth, with the electrophoretic separation of proteins (P. A. Charlwood) and some properties of salt-like crystals (T. R. Bolam) as special cases; colloidal solutions (M. B. M'Ewen); and overpotential (J. O'M. Bockris). Variations in scale and level are perhaps inevitable; the general effect is that the book meets its purpose.

HUMAN BLOOD GROUPS AND INHERITANCE by Sylvia D. Lawler and L. J. Lawler (Heinemann, 1952), pp. 85, 3s 6d.

Described by Dr R. R. Race as 'a readable, accurate and up-to-date account of the human blood groups', this is one of the first

titles in a new 'Scholarship Series in Biology'. The genetics of the rhesus blood groups are cautiously treated: on their nomenclature, although both systems are given, a slurring over of the remaining element of controversy may be doubtfully wise.

VERTEBRATE SEXUAL CYCLES by W. S. Bullough (Methuen's Monographs on Biological Subjects, 1951), pp. 117, 6s.

This is a concise, careful, and well-documented survey of an area of vertebrate biology in which the author has done original work and in which the need for such treatment must have been widely felt.

INTRODUCTION TO STATISTICAL METHOD by B. C. Brookes and W. F. L. Dick (Heinemann, 1951), pp. 288, 21s.

An introduction to statistics, written primarily for students in the natural and social sciences, this is in fact suitable for any serious reader with mathematical equipment up to about sixth-form standard and a preference for practical illustrations.

A STATUE TO MR TRATTLES by C. L. Boltz (Butterworths Scientific, 1952), pp. 168, 12s 6d.

The author's description of his book as 'a collection of pieces on scientific topics' does less than justice to his account of the Trattles case (on the testing of colour vision), which is no less a contribution to history from being made intelligible to the general reader. Other subjects are spread widely; colloids, quicksands, and electrophonic organs are an almost random selection.

THE ELECTRIC CURRENT by P. Dunsheath (Bell, 1951), pp. 211, 18s 6d.

The Christmas Lectures at the Royal Institution for 1949–50, rendered into book form, and generously illustrated.

A DICTIONARY OF PSYCHOLOGY by James Drever (Penguin Books 1952), pp. 316, 3s 6d.

Posthumously published, Professor Drever's dictionary of psychology includes for convenience a number of medical and physical terms encountered in psychological reading.