The Scientific Endeavor

The ocean, the atmosphere, outer space, belong not to one nation or one ideology, but to all mankind, and as science carries out its tasks in the years ahead, it must enlist all its own disciplines, all nations prepared for the scientific quest, and all men capable of sympathizing with the scientific impulse.

JOHN F. KENNEDY

President of the United States of America

CONTRIBUTORS TO THIS VOLUME

MELVIN CALVIN	ROBERT OPPENHEIMER
GEOFFREY F. CHEW	GEORGE E. PALADE
THEODOSIUS DOBZHANSKY	LINUS PAULING
J. B. FISK	I. I. RABI
WILLIAM A. FOWLER	ROGER REVELLE
JESSE L. GREENSTEIN	T. M. SONNEBORN
H. H. HESS	E. L. TATUM
G. EVELYN HUTCHINSON	GEORGE WALD
JOHN F. KENNEDY	VICTOR F. WEISSKOPF
GEORGE B. KISTIAKOWSKY	FRED L. WHIPPLE
ERNST MAYR	JEROME B. WIESNER
NEAL E. MILLER	EUGENE P. WIGNER

The Scientific Endeavor

CENTENNIAL CELEBRATION

OF THE

NATIONAL ACADEMY

OF SCIENCES

THE ROCKEFELLER UNIVERSITY PRESS

LIBRARY OF CONGRESS
CATALOGUE CARD NUMBER 65-18302
PRINTED IN THE
UNITED STATES OF AMERICA
SECOND PRINTING

Foreword

IN THE EARLY DAYS of the nation, our people had a lively interest in science. It could scarcely have been otherwise, for an unexplored continent was a powerful stimulus to the curiosity of adventurous settlers. Even their survival depended upon an understanding of nature. And they came from Europe in a time when a new age of scientific endeavor was being fostered by the young academies of Italy, England, and France: the Accademia del Cimento, the Royal Society of London, the Académie des Sciences.

As our ancestral pioneers cultivated their wild natural environment for the needs of a more social life, they organized societies for discussion of their scientific observations and experiments. The American Philosophical Society Held at Philadelphia for Promoting Useful Knowledge was well established before the Revolution, and the American Academy of Arts and Sciences was founded in Boston soon thereafter. Benjamin Franklin, Benjamin Thompson, and Thomas Jefferson were among scores of colonial Americans who studied nature for pleasure and in the quest of useful knowledge. From them and from succeeding generations of intellectually restless immigrants, there spread a widening interest in science. Literate farmers and artisans and tradesmen gathered in scores of local societies for the discussion of science.

By the middle of the nineteenth century, conditions in the nation had profoundly altered the status of science. The growing acceptance of natural science as a proper subject of instruction in the classical colleges of liberal arts provided a livelihood for increasing numbers of young men who wished to devote all of their time to scientific study, teaching, and research. The Morrill Acts of 1862 made available grants of federal lands for the support of colleges of agriculture and the mechanic arts and thus emphasized the need for wider diffusion of scientific knowledge as the essential foundation for development of our national economy. The Civil War revealed the need of our federal government for scientific and technical advice. Accordingly, 1863 was a propitious time for the Congress to call upon fifty scientists to found the National Academy of Sciences for the encouragement of research and the diffusion of scientific knowledge, and to advise the federal government on matters of science and technology.

During the next half century, the American scientific endeavor increased slowly. Promising young scientists still went to Europe for advanced study and returned to poorly equipped laboratories and heavy teaching schedules that left them little time for research. The members of our federal government were more interested in immediately useful applications of science than in research of greater future value. Our infant industries only reluctantly established laboratories for scientific investigations. The limited scope of Academy activities during those years reflected the nation's preoccupation with the material wealth our burgeoning urban civilization could quickly derive from our vast natural resources.

The second half-century of the Academy began just before the first World War. Again the Academy was called upon to satisfy the urgent scientific and technical needs of the armed forces. This inspired the creation within the Academy of a National Research Council of representatives of the leading scientific and technical societies, which were thus enabled to participate in providing advice to the government, in international scientific activities, and in the general furtherance of science. The growth of universities and the multiplication of industrial laboratories stimulated an ever-increasing demand for scientists; to satisfy this need the Academy, supported by The Rockefeller Foundation, initiated and for more than forty years continued to administer a widespread series of national fellowships for the advanced training of thousands of scientists. As the United States emerged from international isolation, the Academy developed closer ties with its sister academies of other countries and represented the scientists of America in the international scientific unions. The narrower specialization of scientists and the fragmentation of science gave the Academy a vital role as a cohesive center for the synthesis of scientific disciplines. To house these widening activities and to symbolize the national significance of the Academy, the Carnegie Corporation of New York provided the funds for a monumental building in the nation's capital.

The Centennial of the National Academy of Sciences was celebrated in the autumn of 1963.

Many of the 650 members of the Academy gathered in Washington throughout four festive days of social events and brilliant scientific discourse. They were honored by the National Government their predecessors and they had served for a century. Representatives of sister academies, learned societies, and universities throughout the world brought greetings which stressed the world-wide scope and unity of the scientific endeavor—especially appropriate to our Academy in which one quarter of the members were born in foreign countries. A score of foundations and corporations announced their generous intention to complete the House of the Academy that was begun forty years before.

The President of the United States and twenty-three distinguished members of the Academy delivered memorable addresses in which the History of the Universe, the Nature of Matter, the Determinants and Evolution of Life, and the spirit of The Scientific Endeavor were described in a score of remarkably relevant and coherent accounts.

> Day after day the mysteries of life were laid bare, and antecedent to life, the structure of matter, and indeed of the universe, were presented in dramatic and fascinating clarity. . . . the juxtaposition of topics did a very great deal to show the essential unity of scientific disciplines however different their techniques. . . . We listened to great wisdom.

said Professor I. I. Rabi in the final discourse.

Because the addresses contained so much wisdom, were presented with clarity, and revealed the unity of science, it was agreed that they should be made available to a larger audience than could assemble for our Centennial. That is the purpose of this book.

DETLEV W. BRONK
Past President of the National Academy of Sciences
Chairman of the Centennial

FREDERICK SEITZ
President of the National Academy of Sciences

Contents

History of the Universe

FIRST SCIENTIFIC SESSION

The Rosette Nebula in the constellation, Monoceros, showing dark clouds. [*preceding photograph*]

Introduction

ROGER REVELLE

THIS MARKS the birthday of a society which is older than most men but younger than many human institutions.

The National Academy is an infant on the time scale of man's first quest for rational understanding of the universe — the Babylonian and Greek times when men began to think they might make a model in their minds that resembled the world. It is young in terms of the new science that began with Galileo and Newton. Yet many of the scientific events that have forever changed our ideas about the world took place after the Academy was born. Within our own lifetimes we are in the midst of the greatest outpouring of human creativity ever seen, an age comparable to the splendid days of Pericles in Athens or of Lorenzo in Florence, but on an earth-wide scale. The Academy Centennial is a stock-taking of this new age — an attempt to show where man's quest for understanding has led him, what levels of unity in diversity he has reached.

In this great assembly we honor the past and seek the future. Our hope is that by looking backward we can glimpse out of the corners of our eyes something of times to come, the new discoveries that might be made during the next hundred years, and the possible biological and physical events during the next aeons.

The first scientific session is devoted to history, but what a history! It deals with a time span a hundred million-fold greater than the lifetime of our Academy, with horizons of space from which men are forever barred, and with depths of time that are physiologically and emotionally inconceivable. We are concerned with the history of matter in our observable universe — the ways in which simple and uniform particles became organized in diverse elements, large bodies, and complex forms, and finally into men, forms of matter so highly organized and so complex that they can arrive at conclusions about the past and forecasts of the future. We do not deal with the ultimate nature of things, with first beginnings or final ends, but only with the sequences of events that led to the distribution of matter and energy we can see today.

ROGER REVELLE Scripps Institution of Oceanography, San Diego, California

The stories told herein are highly circumstantial, the subjects discussed were hardly respectable a few decades ago. Yet the picture these papers unfold has a majestic consistency. It is based on the assumption that the laws of physics have no history, though the physical constants may have changed with time. This is the faith of science, that physical laws, the relationships between things, are universal, valid everywhere and at all times.

The Greeks had only four elements. It was easy for them to believe that the atoms of these elements were the ultimate things, unchanging and unchangeable. Today we are aware of over 100 elements and more than 1,000 nuclear species. We can no longer think of atoms as ultimate particles, instead we believe that all atoms no matter how diverse are made up of a few kinds of building blocks.

The building blocks of atomic nucleii are neutrons and protons; and the history of the elements is the story of how these "nucleons" became joined together to form the varieties and proportions of elements and isotopes that now exist in the universe. In the first paper, Professor Fowler shows that it is possible to start with a slowly rotating mass of turbulent hydrogen gas, containing only protons and electrons, and through gravitational condensations of part of this matter to form the fiery furnaces called stars. Nuclear burning takes place in the stars, and elements heavier than hydrogen are synthesized.

As the nuclear fuel in a star becomes exhausted, further gravitational collapse occurs, the interior becomes enormously hot, and finally the star suddenly explodes. In this process the heaviest elements are synthesized and both heavy and light elements become widely dispersed in the interstellar gas (stars eject matter into space slowly and continuously, as well as in explosions). A new generation of stars is formed from this more complex material, and the process of synthesis continues. Our sun, though it is billions of years old, is one of the second-generation stars in our own Milky Way galaxy.

As Professor Greenstein points out in the second paper of the symposium, the principal irreversible change in the history of a galaxy is the slow evolution of the interstellar gas, the continual increase in its content of helium and metals, and the decrease in the amount of hydrogen. The loss of nuclear fuel through element synthesis results in a slow decrease in the average luminosity of the galaxy by about 8 per cent per billion years. Within a galaxy, individual stars have widely varying life histories. Some live only a million years; others were born

10 to 15 billion years ago and have hardly changed since. They could continue to exist with little change for thousands of billions of years.

The formation of the elements and the evolution of stars are inseparable parts of the same history. Gravitational forces in an inhomogeneous medium first concentrated some of the primitive galactic material, and the transformation of gravitational potential energy produced the high temperatures required for nuclear processes. But today in our own galaxy nearly all the energy of stellar radiation, and of the explosive processes by which stars finally die and return to the void, is produced in the fusion of light elements into heavier ones, from hydrogen to iron. (Atoms heavier than iron consume energy when they are formed, but their abundances are small.) Elements other than hydrogen could not form, and the stars would never light up, if the properties of the galactic gas were such that gravitational condensation took place in relatively small masses, less than a tenth of the mass of the sun.

When we look outside our own galaxy, we enter a time machine somewhat like the one described by H. G. Wells. The light coming to us from great distances brings us a picture of events that occurred long ago. In the case of the most distant objects, visible through the Mt. Palomar telescope, we are looking backward over 2 to 4 billion years. Here we are beginning to glimpse ancient events that could not have been dreamed of in our scientific philosophy — explosions on an unbelievable scale. The energy emitted in a few hundred thousand years was equal to a hundred million times the total nuclear energy of the sun over its entire lifetime of some ten billion years. These objects apparently had a mass equal to many million suns and a diameter less than one-thousandth that of the Milky Way, yet they gave out twenty times as much light as any galaxy now seen in the heavens. Their luminosity was several hundred billion times the luminosity of the sun. The only source of energy so far conceived of for these grotesque happenings is gravitational collapse on an almost galactic scale. It may be that such explosions were typical of the early history of many galaxies; if so, as Fowler points out, they could account for the characteristic galactic ratio of hydrogen to helium nuclei of about ten to one.

For many scientific generations, one of the most puzzling facts about our own sun and its family of planets was the unequal distribution of mass and angular momentum. The sun contains 700 times the mass of all the planets combined, yet it possesses only 0.5 per cent of the angular momentum in the solar system. It rotates around its axis only once in 25 days. Ever since the days of Kant and Laplace, theories of planetary origin have foundered on this difference in the distribution of mass and momentum between the sun and the planets. Yet all old stars like the sun rotate very slowly, while many young stars rotate with periods of a few hours or days. Such young stars often have strong magnetic fields, whereas the magnetic fields of the sun and other slowly rotating stars are weak. Professor Whipple discusses this problem in his paper on the history of the solar system. He shows that the rising temperature and increasing speed of rotation of the sun as it condensed from a nebular cloud must have produced strong magnetic fields in the disc of plasma, or ionized gas, that surrounded the newborn sun. Angular momentum was transferred along spiralling lines of magnetic force as the plasma particles moved outward, and was acquired by the small solid bodies that later aggregated into planets. Whipple believes these magnetohydrodynamic processes occurred when the plasma disc surrounding the central solar mass had a diameter greater than that of the present orbit of Jupiter.

As Dr. Greenstein states, the differences in speed of rotation between young and old stars throughout our galaxy probably represent a similar magnetohydrodynamic transfer of angular momentum in the processes of planet formation. This is one of the chief reasons why astronomers believe that a considerable percentage of the stars in our galaxy may have planetary systems. Professor Wald in his paper on the origin of life has drawn the obvious implication that life may have originated and evolved at many places in the galaxy.

Wald shows that organisms are composed almost entirely of some 16 to 21 elements. The four elements, hydrogen, oxygen, nitrogen, and carbon make up about 99 per cent of living substance, but such elements as phosphorus, sulfur, and iron play an essential role. Their ability to exchange electrons enables them to be used by living creatures to store, transfer, and release energy. Hence life could not have arisen in the first stages of galactic evolution, when only hydrogen and perhaps helium existed. It was necessary first to form stars in which several stages of nuclear fusion could occur. The elements so created then had to be spewed out into the interstellar gas and reformed into second or later generation stars. Many of these later stars may have evolved planetary systems. But for highly organized life to develop in such a system, at least one planet had to be formed at a distance from the star such that its surface temperature would permit liquid water to exist. At this temperature, the planet would probably be stable for the long period required for organic evolution only if it consisted primarily of silicon, and the other heavier elements, plus oxygen. It would need to be large enough to retain sufficient light elements to provide an atmosphere and liquid water on its surface.

The advent of free oxygen in the earth's atmosphere allowed living things to develop respiration and thereby to obtain energy in high concentrations. A portion of the free oxygen in the upper atmosphere was converted to ozone which shields the earth's surface from destructive ultraviolet radiation. Warm-blooded land animals, producing 10,000 times as much energy per gram of body weight as the sun, could not have evolved before respiration was invented and the ozone layer appeared. Only then could man evolve—that tool-making, time-binding, star-questioning animal. In him, and almost certainly in other creatures like him throughout the starry universe, a new age begins, the age when matter starts to understand its past and consciously to shape its future.

*

The Origin of the Elements

WILLIAM A. FOWLER

> They [atoms] move in the void and catching each other up jostle together, and some recoil in any direction that may chance, and others become entangled with one another in various degrees according to the symmetry of their shapes and sizes and positions and order and they remain together and thus the coming into being of composite things is effected.
>
> Simplicius on Leucippus (A. 14 in Diels, H., and Kranz, W., *Die Fragmente der Vorsokratiker*)

IT IS MY privilege to begin our consideration of the history of the universe with a discussion of the origin of the elements of which the matter of the universe is constituted. The question of the origin of the elements and their numerous isotopes is the modern expression of one of the most ancient problems in science. The early Greeks thought that all matter consisted of the four simple substances — air, earth, fire, and water and they, too, sought to know the ultimate origin of what for them were the elementary forms of matter. They also speculated that matter consists of very small, indivisible, indestructible, and uncreatable atoms. They were wrong in detail but their concepts of atoms and elements and their quest for origins persist in our science today.

When our Academy was founded one century ago the chemist had shown that the elements were immutable under all chemical and physical transformations known at the time. The alchemist was self-deluded or was an out-and-out charlatan. Matter was atomic, and absolute immutability characterized each atomic species. The periodic system of these immutable elements was proclaimed in 1868 by Mendeleev, when the Academy was five years old. Any theory of the origin of the elements was required to account for the formation of each elementary species which remained immutable and unchanged thereafter. An English physician, William Prout, noted that most atomic weights seemed to be integral multiples of that of hydrogen and suggested that all the heavier elements consisted of the lightest one — hydrogen. However,

WILLIAM A. FOWLER W. K. Kellogg Radiation Laboratory, California Institute of Technology

this was a heretical idea and Prout did not sign the two articles on this suggestion which he published in the *Annals of Philosophy,* 1815–1816. He did sign his concurrent articles on the sap of the vine, the excrement of the boa constrictor, the liquor amnii of the cow, and the ink of the cuttlefish. Like many others he is remembered principally for his heresy and not for his orthodox medical and scientific studies.

The Academy was still young when the twentieth century changed all this. In the first years of this century Lord Rutherford and his contemporaries showed that the naturally radioactive elements spontaneously broke down through long chains of intermediate elements to lead or bismuth with the emission of helium nuclei (alpha particles), electrons (beta particles), and photons (gamma rays). We now know that antineutrinos are also emitted in these transformations. Rutherford and Bohr showed that the atom consisted of a central nucleus surrounded by satellite electrons in quantized orbits. The chemical properties of an atom depended in high approximation only on the number of satellite electrons; the radioactive transformations were primarily nuclear. A neutral atom of one element changed into that of another when its nucleus changed its positive charge, and the electronic structure made a secondary adjustment in gaining or losing negatively charged electrons. In 1911 the identification of isotopes by Soddy led to the revival of Prout's hypothesis. Those elements which did not have atomic weights approximately equal to an integral multiple of that of hydrogen were shown to consist of mixtures of isotopes which did have this property. By 1919 Rutherford found it possible to induce nuclear transmutations using the energetic alpha particles from natural radioactive sources. Then in the period 1932–1934 came the deluge of experimental results which serve as the foundation for our current ideas of nuclear structure and the origin of the nuclear species. Urey discovered the deuteron, Chadwick discovered the neutron, and neutrons were soon produced by accelerated deutrons in the laboratory by Crane, Lauritsen, and Soltan. Anderson discovered the positron (antielectron) and the Joliot-Curies produced "artificial" radioactivity in the laboratory. Pauli suggested the idea of the neutrino and the antineutrino and Fermi worked out the consequences of this suggestion in the beta decay processes.

It was immediately realized that the deuteron was the simplest nuclear "molecule" — the deuteron consists of a proton and neutron bound together by the attractive nuclear forces between them. Moreover, all nuclei consist of varying numbers of protons and neutrons, and in this context protons and neutrons soon came to be called nucleons. The neutron was so named because it was found to be electrically neutral. In addition the neutron was found to have approximately the same mass as the proton. Thus the charge number of a nucleus is determined only by the number of protons it contains, while the mass number is determined by the total number of nucleons, neutrons as well as protons, which it contains.

The second simplification which sprang from the discoveries made during 1932–1934 concerned the natural and artificial transformations known as the beta decays. The simplest of these decays is that of the neutron which transforms with a half-life of 11.7 minutes to a proton with the emission of an electron and an antineutrino. The neutron is more massive than its decay products and this excess mass multiplied by c^2 appears as the kinetic energy of these products. In a radioactive nucleus a neutron can also decay to a proton and, moreover, a proton in a higher nuclear energy state than an unfilled neutron state can decay to a neutron. In this case the decay involves the emission of a positron and a neutrino or the capture of an atomic electron with the emission of a neutrino.

It was, of course, not until 1956 with the fall of the conservation of parity that it became clear that antineutrinos really differ from neutrinos in spite of the fact that both have zero mass, charge, and magnetic moment and both have spin one-half or intrinsic angular momentum equal to $\frac{1}{2}\,\hbar$. Antineutrinos move with the velocity of light with their spin vector parallel to their direction of motion as given by the *right-handed* rule. On the other hand neutrinos move with the velocity of light with their spin vector antiparallel to their direction of motion as given by the *left-handed* rule. Positrons, like antineutrinos, are right-handed in beta decay; electrons, like neutrinos, are left-handed in beta decay. It is one of the fundamental properties of the beta decay interaction that *handedness* is conserved. If two leptons, as these light particles are collectively known, are emitted in a beta decay, one must be right-handed, the other left-handed. If one is absorbed and one is emitted as in electron capture, they must have the same handedness.

The essential point is that transformations exist by which a nucleon can change from one form to the other — neutron to charged proton or proton to uncharged neutron. This supplements the nuclear transformations in which neutrons or protons or heavier nuclei can be removed or added to a given nucleus. Neutrons and protons are the basic building blocks of nuclei and under appropriate circumstances they can interchange their nucleonic roles.

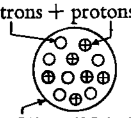
Experiments in high-energy nuclear physics now show that the neutron and proton have internal structure and that they can be annihilated by interactions with antineutrons and antiprotons. Furthermore modern theories in high energy physics treat the neutron and proton as only two of a great number of strongly interacting particles. At the present time it does not seem necessary to postulate astrophysical circumstances in which excitation energies are so great that these considerations are relevant. We start with neutrons and/or protons and ask how the heavier nuclei species have been synthesized at temperatures up to at most $T \backsim 10^{10}$ degrees or interaction energies up to at most $kT \backsim 1$ Mev.

The great simplification which resulted from this concept of neutrons and protons as the fundamental building blocks of nuclei is illustrated most straightforwardly by a consideration of the number of stable and radioactive nuclear species now known, as indicated in Table I. Ninety elements are found terrestrially

Table I

Elements		Isotopes	
Stable	81	Stable	280
Technetium (Stars)	1	Nat. radioactivity	66
Promethium	1	*Art. radioactivity*	1,095
Through bismuth	83	Total	1,441
Nat. radioactivity	9	December, 1961	
Through uranium	92		
Art. radioactivity	11		
Through lawrencium	103		
Neutron	1		
Total	104		

Nuclei = neutrons + protons



Example: $C^{12} = 6N + 6P$

No stable mass 5 or mass 8.

and one more, technetium, is observed in stars; only promethium has not been found in nature. Some 280 stable and 66 naturally radioactive isotopes occur on the earth, making a total of 346. In addition, the neutron, technetium, promethium, and the transuranic elements up to number 103, lawrencium, have been produced artificially. The number of radioactive isotopes artificially produced now equals 1,095, and this number is gradually increasing. In regard to atomic mass numbers, all masses from 1 to 238 are found terrestrially with the important exceptions of mass 5 and mass 8. Laboratory processes have extended the radioactive mass numbers beyond 238 to approximately 260. The total number of nuclear species is now 1,441 with about ¼ of this number known to occur in nature and with ¾ having been produced artificially. In the origin of the elements and their isotopes, we shall find that the radioactive forms often play an even more important role than the stable forms to which they decay in nature.

It will be clear that this picture of the structure of nuclei leads quite straightforwardly to an attempt to explain their origin by a synthesis, or buildup, starting with one or the other or both of the basic building blocks. An alternative which has been suggested is that nuclei now in existence resulted from the breakup of a primordial nuclear fluid with fission and evaporation processes playing a leading role. This point of view has not been quantitatively elaborated in the light of recent evidence on the details of element abundances nor has it been checked in any detail against astronomical observations. The synthesis point of view starts with protons and/or neutrons but does not attempt an answer to the perhaps even more intriguing problem of the origin of these nucleons. Concerning that aspect of the origin problem we have practically no experimental data except on the creation of nucleon-antinucleon pairs in the high-energy laboratory. Since this creation is inevitably followed by annihilation, it is of little direct application in answering the basic problem — the creation and survival of our immediate astronomical system, the Galaxy, which we know to be composed of particles (protons, neutrons, and electrons) and not antiparticles. There is some evidence that neighboring galaxies in the cluster to which the Galaxy belongs also consist of particles — concerning galaxies outside our cluster we have no information. All of this is to be contrasted to the growing body of nuclear and astrophysical evidence concerning the synthesis of nuclei from nucleons. It is to this problem that the discussion in this paper is directed.

With an acceptance of the existence of nucleons, this question can be asked, "What has been the nuclear history of the matter, on which we can make observations, which produced the elements and their isotopes in the abundance distribution which observation yields?" To attempt to understand the sequence of events leading to the formation of the elements, it is necessary to study the so-called "universal" or "cosmic" abundance curve. Such a curve in schematic form is shown in Figure 1. This figure gives abundances by number of atoms with 10^6 atoms of silicon taken as an arbitrary standard. It is taken from the analysis of abundances made by Suess and Urey in 1956. Abundance curves are derived

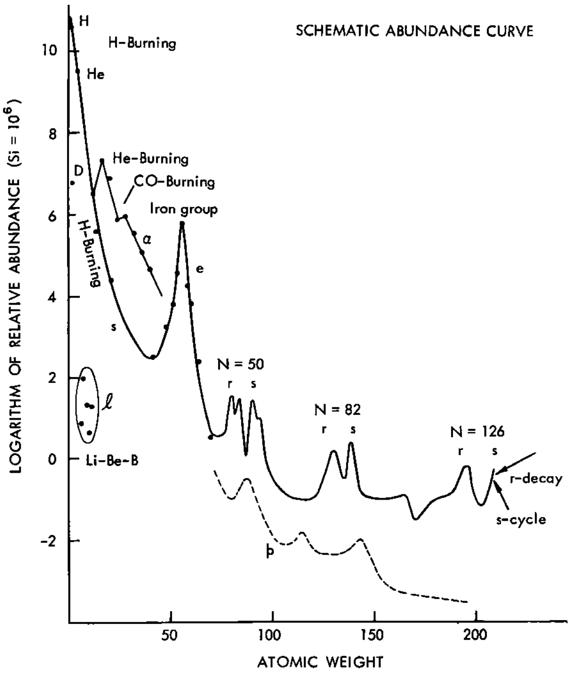


FIGURE 1. Schematic curve of atomic abundances as a function of atomic weight based on the data of Suess and Urey. Suess and Urey have employed relative isotopic abundances to determine the slope and general trend of the curve. There is still considerable spread of the individual abundances about the curve illustrated, but the general features shown are now fairly well established. Note the overabundances relative to their neighbors of the alpha-particle nuclei with $A = 16, 20, \ldots, 40$, the peak at the iron-group nuclei, and the twin peaks at $A = 80$ and 90, at 130 and 138, and at 195 and 208.

mainly from terrestrial, meteoritic, and solar data and, in some cases from other astronomical sources.

Whether or not this abundance curve is truly universal is not of too great relevance. It is the distribution for the great bulk of the matter on which we have been able to make observations. It must be emphasized that it is heavily weighted by observations on matter in the solar system: the sun, the earth, and the meteorites. Some additional information comes from spectroscopic observations on nearby stars and on the gas and dust which lie between them and which scatter and reflect starlight. All in all, with some exceptions, this material has much the same composition. We can seek the history of this particular matter. We can also ask for the history of the peculiar and abnormal abundances observed in some stars. In time, we may obtain more information on the abundances in other parts of our Galaxy and in other galaxies, and only then shall we be in a position to approach the problem of what part of the average abundances is truly universal or cosmic.

The schematic curve shown in Figure 1 is rich in detail, the complete abundance curve even more so. Superimposed on the figure are certain abbreviations (e.g., H-burning, He-burning, CO-burning) and symbols (e.g., α, e, r, s, l) designating nuclear processes which may have been involved in the synthesis of the relevant nuclear species. At our present state of knowledge it is difficult to see how all these processes could have occurred in a single astronomical event such as a primordial explosion or "big bang" at the moment of creation of the universe. Moreover, abundances are not universal as discussed above although there may eventually prove to be some underlying universality in the abundance curves of all astronomical systems. The only single-event theory of element synthesis which has been worked out in detail is that of Gamow, Alpher, and Herman. The restrictions placed on the nuclear processes by the density-temperature conditions assumed for cosmological reasons in this theory lead to an apparently insuperable difficulty at mass 5 at which no stable nuclear form exists. Gamow has phrased the point in the following words:

> . . . . However, since the absence of any stable nucleus of atomic weight 5 makes it improbable that the heavier elements could have been produced in the first half hour in the abundances now observed, I would agree that the lion's share of the heavy elements may well have been formed later in the hot interiors of stars.[1]

while Salpeter has written:

> . . . . Thus, for building all elements heavier than helium, the original expansion of the universe is, from the nuclear point of view, simply useless.[2]

In any case, the major developments in recent years have followed the hypothesis that the elements heavier than hydrogen have been synthesized in stars where the varying conditions in stars of different mass and the varying conditions in a given star as it ages and evolves lead to a plethora of circumstances under which nucleosynthesis can take place. Atkinson, Houtermans, Sterne, von

Weizsäcker, Bethe, Critchfield, and others made important contributions in early studies. Hoyle, Bondi, and Gold accepted stellar nucleosynthesis as a necessity in their steady-state cosmology. The details of many nuclear processes which may occur in stars have been worked out by Salpeter, Lauritsen, Greenstein, and others, and a general account including several new processes has been given by Burbidge, Burbidge, Fowler, and Hoyle. Several groups in Japan and in Russia have made major contributions in this field. The real heroes are the many experimentalists who have painstakingly measured nuclear reaction rates at the excruciatingly low levels which occur in the laboratory when appropriate astrophysical energies ($kT \backsim 1$ to 300 kev) are approached and the many observationalists who make difficult spectroscopic studies of element abundances in distant, faint stars. No less heroes are those who measure minuscule abundances of the heavy elements in rocks, meteorites and tektites.

In what follows the origin of the elements will be discussed in the context of stellar nucleosynthesis and the illustrative examples will be those most familiar to the author. Before continuing it is well to emphasize certain points, the chief of which is this: *stellar nucleosynthesis can be incorporated into almost all the cosmologies which have been studied to date.* In the evolutionary cosmologies which adopt a finite age for the universe it is necessary in the beginning only to create nucleons but not nuclei. If the red shift observations are taken to indicate an expanding universe, then under the relatively high densities and temperatures at the beginning, the neutron was the predominant nucleon. With time and expansion the neutrons decayed to protons and electrons which could form neutral hydrogen atoms which in turn could aggregate into galaxies and stars because of gravitational forces. Primordial nucleosynthesis need not be necessarily assumed but on the other hand some such nucleosynthesis may have occurred. For example some helium may have been produced in the beginning. Synthesis up to and including mass 4 but not beyond meets no insuperable problems. The astronomical evidence does not exclude the possibility that the Galaxy formed with some helium as well as hydrogen.

In the evolutionary cosmologies which assign an infinite age to the universe by adopting an appropriate value of the cosmological constant, the early form of matter can be taken to be hydrogen which again ultimately formed astronomical systems. In the steady-state cosmology the steady creation is that of neutrons or of protons and electrons (to conserve charge). This is at least the case in our corner of the universe. If antinucleons are being simultaneously created now along with nucleons, the resulting annihilation would in part produce neutral pions which decay with the emission of ∽100 Mev gamma radiation. Space probes have not found evidence for such radiation and place a low upper limit on the creation rate for antinucleons in our astronomical neighborhood.

The separate creation of nucleons and leptons now or at some remote time in one or the other cosmological circumstance violates present experimental findings. In the laboratory nucleons can only be created or annihilated with antinucleons, leptons can only be created or annihilated with antileptons. The problem is common to all cosmologies. There is little inkling of the solution and we must content ourselves with some knowledge of *synthesis* but with little or none concerning *genesis*.

To set the stage for an exposition of current ideas on stellar nucleosynthesis, it is necessary to give some account of the origin and history of the Galaxy and of the solar system. This will be done in succeeding papers by Professors Greenstein and Whipple. Here we give only a brief resume. When we look out beyond the confines of the planetary system which surrounds the sun, we see our Galaxy as a majestic assemblage of stars which we call the Milky Way. From our position approximately half-way out from its center the system is viewed edge on. The stars in the Galaxy populate a flat, disk-shaped structure with a spherical nucleus from which radiate several spiral arms trailing the direction of rotation. A few old, high-velocity stars occupy a spherical "halo" above and below the equatorial disk. Twelve to 15 billion years ago the Galaxy was vastly different. At that time according to current ideas, it was a rotating mass of turbulent hydrogen gas. Because of statistical fluctuations, the gas in regions of relatively low turbulence and high density condensed into stars under the influence of gravitational forces. As the stellar material contracted, the interior became very hot and dense from the conversion of gravitational potential energy into thermal kinetic energy. These conditions serve to "trigger" exothermic nuclear reactions, beginning with the fusion of hydrogen into helium and going on, as we shall see, into more complicated processes. The energy released from these fusion reactions makes them self-sustaining until the nuclear fuel is exhausted. In addition, this energy release leads to the development of internal pressure which stabilizes the star against further gravitational collapse. This stability lasts for the relatively long periods necessary to consume the interacting nuclei. The burning hydrogen in the sun has lasted for 4.5 billion years and will last as long in the future. We emphasize once again that this is nuclear, not chemical, burning.

A self-sustaining fusion process has not been successfully accomplished terrestrially in spite of valiant efforts to do so in many countries. In stars the containment problem, which is so difficult to solve on a terrestrial scale, has been solved automatically by the large mass of these celestial objects. Stars can thus be considered as gravitationally stabilized fusion reactors which release energy through the conversion of one form of nuclear matter into another. Gravitational energy can raise the temperature to the "ignition point" for nuclear processes, but it cannot serve as the source of energy in a stable star which is no longer contracting; the nuclear reactions themselves do this.

As successive nuclear processes take place, the composition of a star changes and the star is said to evolve as its internal structure and external appearance vary in response to these composition changes. It is essential in the point of view of stellar synthesis that instabilities arise during the evolution and aging of a star that return the transmuted material to interstellar space. It is there mixed with the uncondensed hydrogen gas in the Galaxy so that it is available for condensation into second- and later-generation stars. The general state of affairs in this "equilibrium" between stars and the interstellar gas and dust is illustrated in Figure 2.

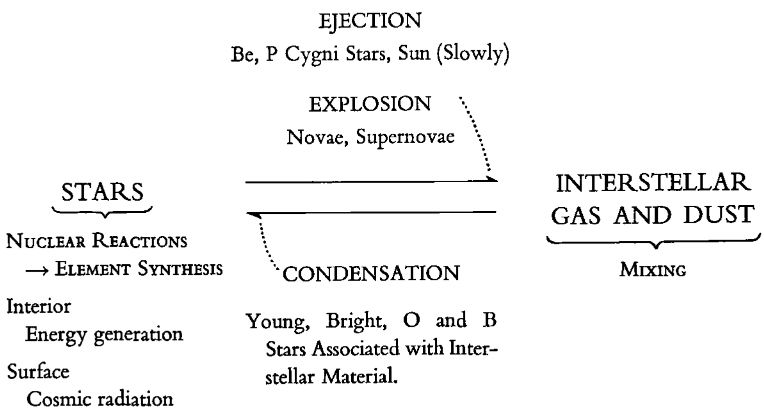


FIGURE 2. Transfer of material between stars and interstellar gas and dust. Synthesis of elements occurs in the stars, and mixing to yield the relative abundance of the elements occurs in interstellar space. Mechanisms for the transfer as observed astronomically are indicated.

Stellar nucleosynthesis demands that there exist this interchange of material between stars and the interstellar medium of gas and dust. The stars are the nuclear furnaces; the space between is the site of the mixing and dilution which result in the average abundance distribution over fairly large astronomical regions. Observations confirm that matter is given off by stars, both slowly and explosively, and that new stars are continually forming from the interstellar material. Giant stars lose mass at a fairly substantial rate; even our sun slowly ejects matter into space. The planetary nebulae, such as that shown in Figure 3, show spherical "smoke rings" moving away from a central star. The "rings" are actually shells. The most spectacular instabilities in stars result in the novae and supernovae that are observed to flare up suddenly and then die away in brightness. For novae a mass loss of the order of 0.1–1 per cent can suddenly occur. In supernovae explosions, all or a substantial fraction of the mass of a star may be ejected with a high velocity, 10^3 to 5×10^3 km/sec., into space. Such an explosion results after years of expansion in an amorphous mass of material such as the Crab nebula (Figure 4), which is now located in the same region in the sky where Chinese astronomers observed the appearance of a "guest star" in A.D. 1054. Quantitative calculations show that the rate of these mechanisms is such that the heavy-element abundance in the solar system could have been synthesized in earlier stars, which formed and evolved in the Galaxy, if the Galaxy is several billion years older than the sun, which seems to be the case from other evidence.

The process, the reverse of the breakup of stars, the formation of new stars, also has substantial observational confirmation, albeit somewhat indirect. There are stars in the heavens so bright for their known mass that even nuclear processes cannot have kept them shining for more than a few million years. They are thus much younger than the sun and the Galaxy. The bright stars are "young stars," and since they occur only in regions observed to be populated with relatively large amounts of gas, it is reasonable to assume that they condensed from the interstellar material. In regions of no gas and dust there are few great bright stars; only old stars of low mass, low central temperatures, and low nuclear reaction rates are found there. They are the slow burners from the original and succeeding condensations that cleaned up the vicinity in which they are located.

The process of gravitational contraction of a protostar containing only hydrogen leads to a temperature and density rise in the interior and to a gradient in these quantities such that the temperature and density are highest at the center

FIGURE 3. The "ring" planetary nebula in Lyra showing the spherical shell of gas which is moving away from the central star and was presumably ejected by it. The off-center star within the ring is a field star.

FIGURE 4. The Crab nebula photographed in the wave length range λ6300 to λ6750. The filamentary structure stands out clearly at this wave length, which comprises light mainly due to the Hα line. The nebula consists of the expanding debris of a supernova which was observed to occur at the same point in the sky by the Chinese in A.D. 1054.

of the star and drop off rapidly to relatively low values at the stellar surface. When the central temperature reaches 10^7 degrees absolute and the density reaches 100 gm/cm^3, the hydrogen begins to interact through the so-called direct proton-proton or p-p chain which consists initially of the reactions shown in Figure 5. Each of these reactions is exoergic. The reactions proceed with the emission of large amounts of energy even when the reactants have relatively low energies. The reaction products rapidly lose their large energies by atomic (not nuclear)

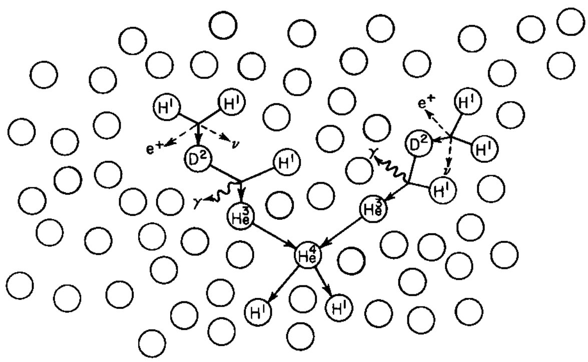


FIGURE 5. Schematic representation of the fusion of hydrogen into helium by the p-p chain which occurs in main sequence stars of one solar mass or less. Density, 10^2 gm/cm^3. Temperature, 10^7 degrees K.

collisions, and thus the reactions are not reversed. The over-all result is $4H^1 \rightarrow He^4$ with the emission of 26.7 Mev or 4.3×10^{-5} ergs of energy from the small fractional difference (0.7 per cent) between the mass of the helium atom and that of four hydrogen atoms. From the standpoint of nucleosynthesis, the p-p chain is important because it is a mechanism by which pure hydrogen can be converted into helium.

The p-p chain is now thought to be the predominant mode of conversion of hydrogen into helium in the sun. In the lifetime of the sun a considerable amount of He^4 has been produced. This He^4 reacts with He^3 and enters catalytically into the production of additional He^4. This can be seen from the complete set of major p-p chain reactions written in modern nuclear notation as follows:

$$H^1(p,\beta^+\nu)D^2(p,\gamma)He^3(He^3,2p)He^4$$

or $$\hookrightarrow (He^4,\gamma)Be^7(e^-,\nu)Li^7(p,\alpha)He^4$$

or $$\hookrightarrow (p,\gamma)B^8(\beta^+\nu)Be^{8*}(\alpha)He^4$$

In the two last alternatives the interaction of He^4 with He^3, with the addition of a proton, finally results in the production of two He^4. It will be noted that neutrinos of various energies are emitted in this chain. Those accompanying the production of D^2 have an average energy equal to 0.25 Mev. Those from the decay of Be^7 fall into two approximately monoenergetic groups at 0.88 Mev (90 per cent) and 0.40 Mev (10 per cent). The neutrinos from B^8 have an average energy equal to 7.4 Mev. The relative production of these various neutrinos depends sensitively on the central temperature of the sun. Their detection cross-sections increase rapidly with energy. Thus a solar neutrino detection experiment such as that under way by Davis at Brookhaven using $Cl^{37} + \nu \rightarrow A^{37} + e^+$ may serve to make an independent determination of the sun's central temperature which can at the present time be inferred only from models of the sun's internal structure. It would seem there is available what might be called a *neutrino thermometer.*

The fusion of protons into helium can occur in stars even though protons are all positively charged and mutually repel each other. As a matter of fact, on classical Newtonian mechanics, the fusion cannot occur, because even at stellar temperatures the protons do not have sufficient relative velocities to overcome their mutual repulsion. Sir Arthur Eddington, who proposed hydrogen fusion as the source of energy in stars in 1920, gave a magnificent answer to those who criticized him on classical grounds: "... We do not argue with the critic who urges that the stars are not hot enough for this process; we tell him to go and find a hotter place."[3] Eddington's critics were saved from their classical fate by modern quantum mechanics, which governs the behavior of atomic particles and permits fusion to occur even when it is "impossible" on Newtonian mechanics.

Stars which live and shine from energy generated through the process $4H^1 \rightarrow He^4$ fall in a luminosity-color classification called the "main sequence." However, as the hydrogen in the central regions of the star is exhausted, the star ceases to be homogeneous in composition throughout its interior and will move, or "evolve," off the main sequence. The conversion of hydrogen "fuel" into helium "ash" occurs in the core of the star because the temperature and density are highest there. Judging from astrophysical observations, it appears that the reaction product helium is mixed with the outer envelope, still hydrogen, with extreme difficulty. Thus, a core of helium develops and gradually increases in size as more and more hydrogen is converted. Because of greater electrostatic repulsions, the doubly charged He^4 does not burn at 10^7 degrees or even at considerably higher temperatures, and so energy generation ceases except in a thin shell surrounding the helium core. This shell now contains the hottest hydrogen in the star. It has been estimated that the shell temperatures reach 3×10^7 degrees, while the density is of the order of 10 gm/cm^3. In the central regions, the nuclear hydrogen furnace goes out for lack of fuel, and one would expect from ordinary experience with furnaces that the temperature would drop. But this is not at all the case in stars because of their great potential gravitational energy. The helium in the core begins to contract and its temperature rises as gravitational energy is converted into kinetic energy.

This "anomalous" behavior of stars is not all pure conjecture, for the sudden rise in temperature of the core also heats up the envelope, which expands enormously and increases the surface area of the star. The increased area means that energy can be radiated at a lower surface temperature, and thus the surface reddens in color. Larger in area and redder in color than main-sequence stars of the same luminosity, these stars are aptly called the "red giants" by astronomers.

Eventually the helium in the core reaches temperatures ($\backsim 10^8$ degrees) and densities ($\backsim 10^5$ gm/cm^3) at which Coulomb repulsions should no longer critically inhibit nuclear processes between two helium nuclei. What these processes might be constituted for a long time the Gordian knot of nuclear astrophysics. Two helium nuclei, upon interacting, might be expected to form Be^8. However, as noted previously, no nucleus of mass 8 exists in nature, and from this early investigators inferred that it must be unstable. Shortly after World War II this was confirmed in quantitative measurements of the Be^8 decay at Los Alamos and the California Institute of Technology. In both laboratories it was found that when Be^8 was produced artificially in nuclear reactions, it promptly broke up into two alpha particles. However, the energy of breakup was found to be relatively small, slightly less than 100 kev. With this last fact in mind, Salpeter of Cornell University then pointed out that, although hot interacting helium in a star will not produce a stable Be^8 nucleus, it will produce, at 10^8 degrees and 10^5 gm/cm^3, a small but real concentration of Be^8 as a result of the equilibrium between the formation and breakup processes. Now, nuclei are found in the laboratory to capture alpha particles with the emission of energy in the form of gamma radiation. Salpeter pointed out that the Be^8 should behave similarly and that if, after its formation from two alpha particles, it collided with a third, the well known stable carbon nucleus C^{12} should be formed. Because of the low equilibrium concentration of the Be^8, about 1 part in 10 billion at 100 million degrees, Hoyle emphasized that the Be^8 capture process had better be a very rapid one, or a "resonant" reaction in nuclear parlance. Experiments at Stanford, Brookhaven, and Cal Tech have shown that this is the case. It has been possible to show that there exists an excited state of the C^{12} nucleus at 7.656 Mev, with almost the exact energy of excitation and other properties which Hoyle predicted that it must have in order to serve as a thermal resonance for the formation of C^{12} from Be^8 and He^4 in stars.

Thus there now exists a reasonable experimental basis for the two-stage process by which three alpha particles in the hot dense cores of red giant stars can synthesize carbon, bypassing the intervening elements lithium, beryllium, and boron. This process is indicated schematically in Figure 6. The over-all process can, in fact, be looked upon as an equilibrium between three helium

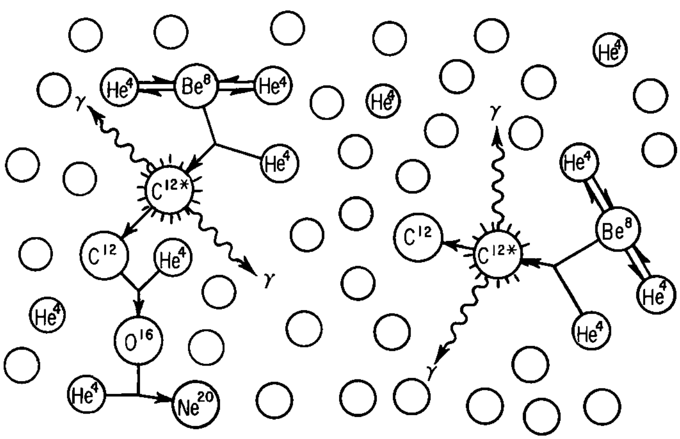


FIGURE 6. Schematic representation of the fusion of helium to form C^{12} which occurs in red giant stars. Density, 10^5 gm/cm^3. Temperature, 1.3×10^8 degrees K.

nuclei and the excited carbon C^{12}, with occasional irreversible leakage out of the equilibrium to the ground state of C^{12}. In reaction notation, we have

$$3He^{4} \rightleftarrows C^{12} \rightarrow C^{12}.$$

The C^{12} frequently captures a helium nucleus to form O^{16} before the helium is exhausted. In extreme cases this results in the over-all process $4He^{4} \rightarrow O^{16}$. In stars, there is no difficulty at mass 5 and the difficulty at mass 8 has been surmounted. When the central conditions in a red giant reach 10^{8} degrees and 10^{5} gm/cm^{3}, the helium begins to burn and energy is released. Because of the small fraction, 0.07 per cent, of mass converted into energy in the above process, the red giant star is not stabilized for any long period after the onset of the helium burning. The major release in nuclear energy comes in the first process, $4H^{1} \rightarrow He^{4}$. In any case, however, the astronomical evidence indicates that the trend toward catastrophic internal temperatures is stopped and the evolutionary track reversed. Stars that become unstable at this point will eject unburnt hydrogen and helium and the synthesized carbon and oxygen into interstellar matter. Others that remain stable will continue the synthesis process.

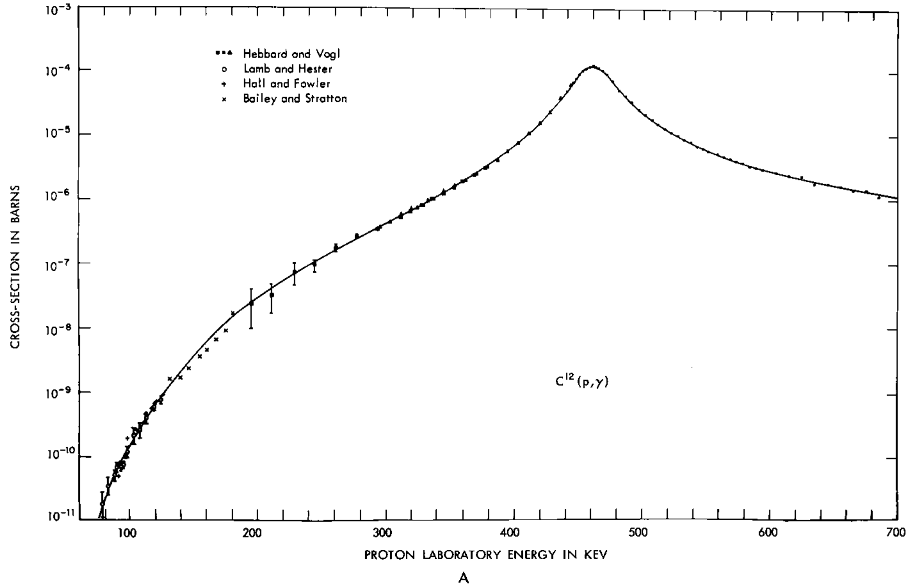
The C^{12} and O^{16} ejected by stars which become unstable will mix with the primordial interstellar matter and eventually condense into a "second" or later generation star. In this new star hydrogen can be converted into helium through what is now called the CNO bi-cycle since it incorporates the original CN cycle of Bethe and von Weizsäcker and a branch involving O^{16} and O^{17}. In modern nuclear notation the reactions are

$$\rightarrow C^{12}\ (p,\gamma)\ N^{13}\ (\beta^{+}\nu)\ C^{13}\ (p,\gamma)\ N^{14}\ (p,\gamma)\ O^{15}\ (\beta^{+}\nu)\ N^{15}\ (p,\alpha)\ C^{12}$$

or

$$\rightarrow N^{14}\ (p,\gamma)\ O^{15}\ (\beta^{+}\nu)\ N^{15}\ (p,\gamma)\ O^{16}\ (p,\gamma)\ F^{17}\ (\beta^{+}\nu)\ O^{17}\ (p,\alpha)\ N^{14}$$

These reactions have been extensively investigated experimentally. Recently measured laboratory cross-sections for C^{12} (p,γ) and C^{13} (p,γ) are shown in Figure 7. The solid curves are theoretical cross-sections fitted to the data by the adjustment of four phenomenological parameters—the resonance energy, the radius of interaction, and the probabilities at resonance for proton absorption and gamma ray emission. Effective thermal energies in hydrogen



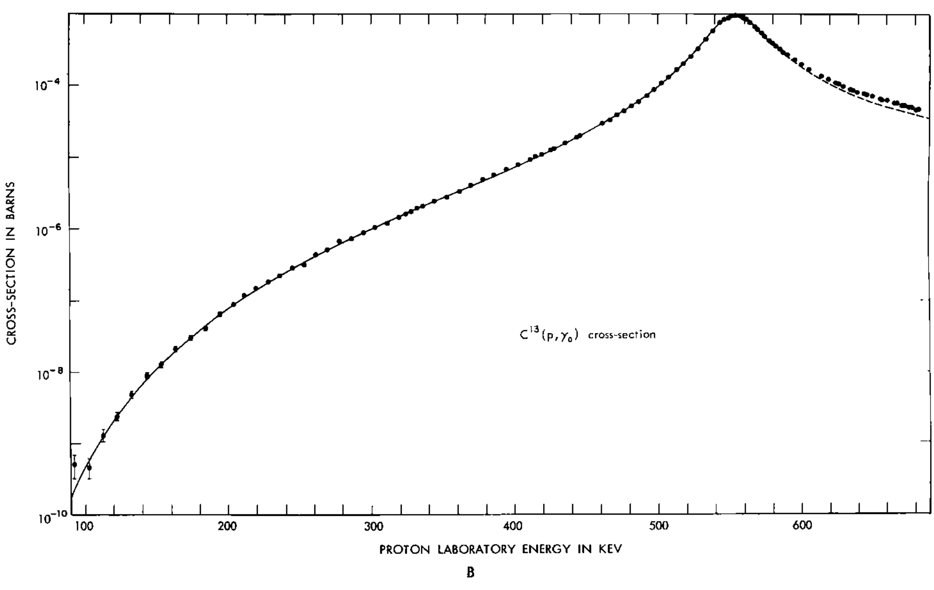


FIGURE 7. The dependence on energy of the cross-section for the reaction (*a*) $C^{12}(p,\gamma)$ and (*b*) $C^{13}(p,\gamma)$. Experimental points are compared with a four-parameter theoretical curve. Since 1946, the Office of Naval Research has supported the experimental work in our laboratory by which these and many other cross-sections have been measured.

burning in stars correspond to 10 to 50 kev and fall below the lowest energies at which the reactions are detectable in the laboratory. Even at 100 kev the cross-sections are only $\backsim 10^{-34}$ cm^2! The excellent agreement with theoretical expectations leads to some confidence in the extrapolation of the data to stellar energies. This is customarily done by dividing the cross-section by the main energy dependence of the Coulomb penetration factor to obtain the so-called cross-section factors illustrated in Figure 8. The cross-section factor can be accurately extrapolated to zero and can then be integrated over a weighting function consisting of the penetration factor and the Maxwell-Boltzmann distribution in thermal energies.

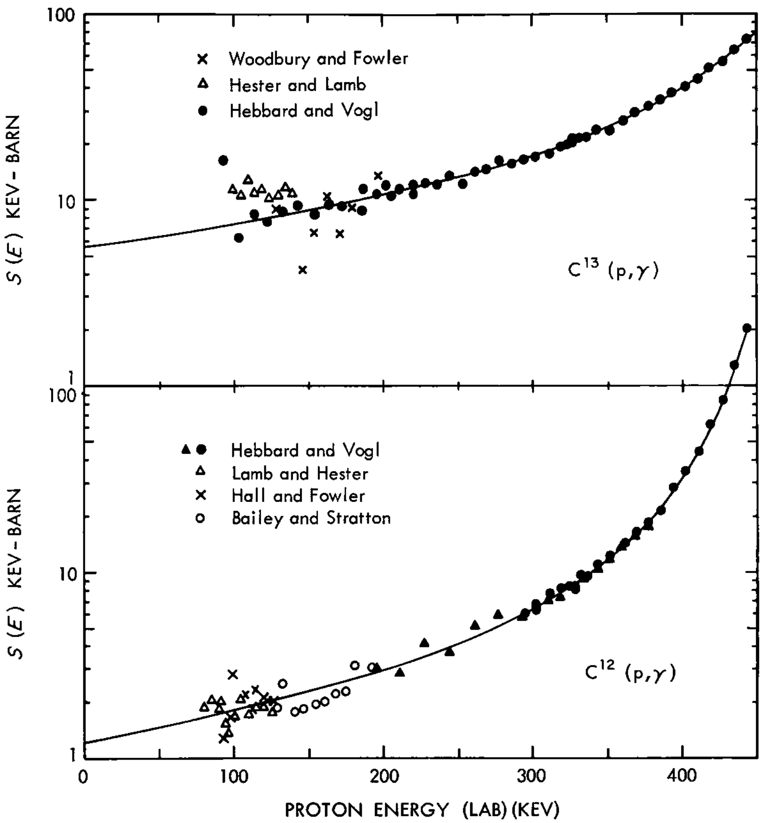


FIGURE 8. The dependence on energy of the cross-section factors for the reactions $C^{12}(p,\gamma)$ and $C^{13}(p,\gamma)$.

If resonances occur below the lowest energies measured, then the thermal cross-sections and reaction rates would be considerably greater than given by the extrapolation of the laboratory data and still not be directly detectable in the laboratory. Fortunately, separate studies involving the compound nuclei, N^{13} and N^{14}, in the cases under consideration can be made to show that no sharp resonances corresponding to excited states in these nuclei contribute to the cross-section in the important but unobservable thermal region.

Returning to helium burning the nuclear evidence indicates that this process should only rarely proceed beyond oxygen. Thus, C^{12}, O^{16}, or a mixture of both are the products of helium burning. Eventually the helium is exhausted; the core of carbon and oxygen contracts and heats up until the C^{12} and O^{16} begin to burn. The result is the production of a number of intermediate mass nuclei among which Ne^{20}, Mg^{24}, Si^{28}, and S^{32} are the most abundant. This is expected from the great nuclear stability of these nuclei with mass number an integral multiple of 4. The great stability is most simply understood in terms of the model in which these nuclei consist of complexes of the highly stable alpha particle. Indeed these nuclei are the most abundant among the isotopes of the elements neon, magnesium, silicon and sulfur. In a star which remains stable the evolutionary process continues. The Coulomb repulsions between nuclei with $Z = 10$ to 16 are very strong and burning no longer proceeds by the simple fusion of the interacting products. Instead as the temperature rises a number of the intermediate nuclei are photodisintegrated in the intense high-energy flux of the tail of the Planck distribution with the emission of alpha particles. For example a Si^{28} nucleus can be broken down into seven alpha particles at temperatures near 3×10^9 degrees. These alpha particles are captured by other nuclei which escaped photodisintegration. Thus another Si^{28} nucleus can capture seven alpha particles to form Ni^{56}. The over-all result is $2Si^{28} \rightarrow Ni^{56}$ but the detailed mechanism is not direct fusion but the alpha-process in which buildup of one nucleus to double its original mass and charge occurs upon the breakdown of another into alpha particles. Ni^{56} is radioactive and decays through Co^{56} to Fe^{56} through the successive capture of two electrons from the plasma continuum in a star with the emission in each capture of a neutrino. Many other nuclei near Fe^{56} from Ti^{46} to Ni^{62} are thought to be produced in this way.

From the standpoint of nuclear physics, it is clear that the sequence of successive burning of heavier and heavier nuclei through charged-particle reactions should terminate at the iron-group nuclei, which are the most "stable" nuclei in the sense that the internal neutron-proton energies are at a minimum and their binding energies are at a maximum in absolute magnitude. Both heavier and lighter nuclei have higher internal energy content and are less stable in this sense than the iron-group nuclei.

Very high temperatures and great densities will be reached at the production of the iron-group elements. Under these conditions the rates of all possible reactions will be very great indeed and the situation will be best described in terms of a nuclear equilibrium. This appears to be indeed the case, because the shape of the iron-group peak illustrated in the abundance curve of Figure 9 has been found by Burbidge, Burbidge, Fowler, and Hoyle to be in good agreement with the calculated equilibrium distribution at 3.8×10^9 degrees and 3×10^6 gm/cm^{-3} and with a free proton-to-neutron ratio of 500:1. The temperature and density are consistent with the conditions leading up to equilib-

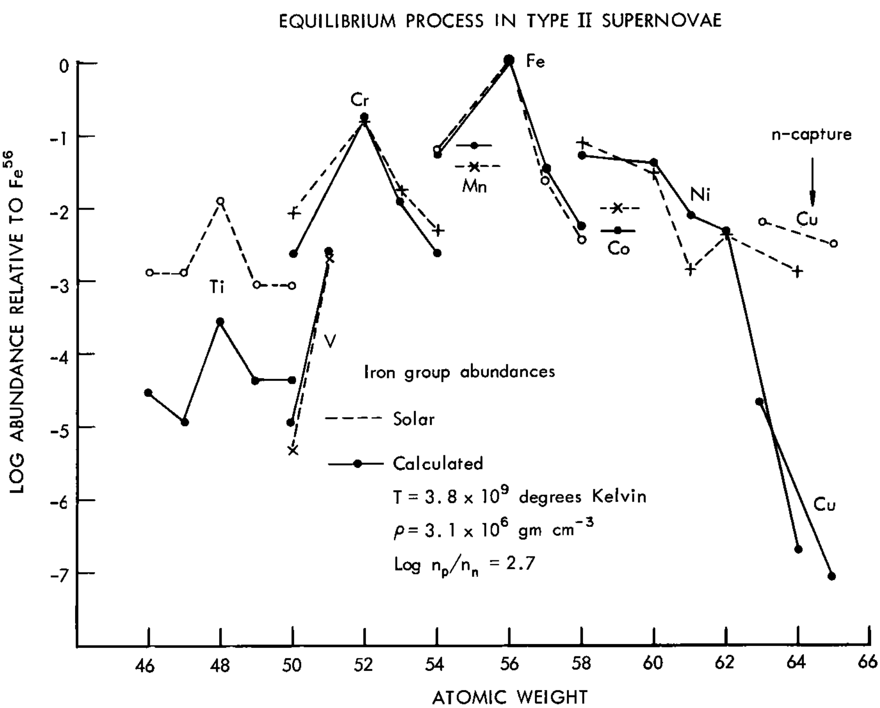


FIGURE 9. The abundance relative to Fe^{56} of nuclei produced in the *e*-process.

rium and the free proton/free neutron ratio is an essential parameter in determining the proton and neutron numbers in the nuclei produced at equilibrium. In the calculation it was necessary to take into account the experimentally known properties of the ground and low-lying excited states of the stable and β-active nuclei involved in the equilibrium and to take into account the β decays on freezing of the mixture. In Figure 1 the process is designated the *e*-process. It is the *e*-process by which the iron-group elements are synthesized. Their overabundance relative to their neighbors can be understood on the basis that enough stars remain stable long enough to develop an iron "ball" in their centers at the end of a long line of energy-generating charged-particle reactions.

The α-process and the *e*-process probably occur at a rapidly evolving or even explosive stage of stellar evolution. It has been suggested that the collapsing core of a star in its terminal stages as a red giant or in its final catastrophic supernova stage is a possible site for such processes. The collapse of the core is brought about by the fact that no further generation of nuclear energy occurs after the iron-group nuclei are produced. Gravitational contraction takes place unimpeded. The implosion is actually speeded up in the inner regions of the core by the refrigerating action of nuclear processes which transfer some of the iron-group nuclei back into lighter nuclei, mostly He^4 and neutrons, with the absorption of energy.

The implosion of the core removes the underlying support of the envelope material of the star, which contains unevolved nuclear fuel capable of releasing large amounts of energy on being raised to high temperatures. The gravitational collapse of the envelope material does just this. The energy release by the nuclear reactions in the envelope material further raises its temperature, the collapse is reversed by expansion of the material, and all or part of the envelope material and probably even a portion of core material are blown out from the star at high velocity. The result is observed astronomically as the occurrence of a supernova in which a star is observed in a very short interval to flare up to many times its previous luminosity and to eject a large fraction of its mass into space.

The time scale of the *e*-process has recently been extensively investigated in the light of the large energy losses to be expected on current beta decay theory from the annihilation of electron-positron pairs according to the reaction

$e^+ + e^- \rightarrow \nu + \bar{\nu}$. At the high temperature of the e-process many electron-positron pairs are produced in the interaction of radiation and nuclei. Equilibrium is established when production and annihilation are equal. The neutrino emission competes with $e^+ + e^- \rightarrow \gamma + \gamma$ in only about one case in 10^{19}, but the gamma rays are trapped in the star whereas the neutrinos and antineutrinos escape directly with the velocity of light. They escape with the kinetic energy and rest mass equivalent energy of the pair and constitute a critical drain on the dwindling energy resources of the stellar interior. The evolutionary process is speeded up and the burning processes from O^{16} to Ni^{56}, which would otherwise require 10^4 years for completion, take place in approximately 1 day. The time available for the typical decay, Ni^{56} to Fe^{56}, in the e-process is even shorter in the range 10^3 to 10^4 sec. The question arises: How does this short time interval affect the resultant equilibrium process abundances? The term "equilibrium" is applicable only in the sense that ordinary nuclear processes involving nucleons, nuclei, and gamma rays are proceeding very rapidly even relative to the short over-all time set by the neutrino loss. On the other hand the slow electron captures proceed in times comparable to the loss time. In Table II, a calculation is made under conditions appropriate to a star with mass equal to 30 solar masses for the final abundances of the isotopes of iron as a function of the time available for electron capture. In the first row under the heading the results for "zero" time at equilibrium are given. At this time the α-process has produced nuclei

Table II. Iron Isotopes—Per Cent of Total E-Process Abundance by Mass

$\overline{Z/N}$	$\log n_p/n_n$	Fe^{54}	Fe^{56}	Fe^{57}	Fe^{58}	Electron capture time (10^4 sec.)
1.000	8.6	1.7	89.1	2.9	0.0	0.0
0.950	6.6	43.4	21.9	7.2	0.0	0.2
0.900	4.0	34.0	29.6	4.7	0.04	1.2
0.872	2.7	4.3	66.6	2.5	0.23	3.2
0.860	1.2	0.2	64.5	3.0	4.0	∞
						$\uparrow$
Solar values		4.1	65.0	1.6	0.23	$M = 30\ M_\odot$

with equal numbers of protons and neutrons so that the average over-all nuclei yield $\bar{Z}/\bar{N} = 1$. It is found that the ratio of free protons to free neutrons in the "gas" surrounding these nuclei is 4×10^8. The most abundant nucleus is Ni^{56} which constitutes 89.1 per cent of the material by mass. If the α-process material were immediately ejected from the star at this juncture, the Ni^{56} would eventually decay to Fe^{56} which would then have an abundance equal to 89.1 per cent of the *e*-process group. The expectations for Fe^{54}, Fe^{57}, and Fe^{58} are also given. These do not agree at all with the observed values for the sun, given in the last row. Moreover, the complete equilibrium values given in the next to the last row do not agree with observations. These values are calculated for a long time compared to characteristic electron capture times and correspond to the free proton/free neutron ratio expected if all the beta decay processes—electron and positron capture and emission—reach equilibrium. Thus the times available for the decays were neither very short nor very long compared to the decay times. As indicated in Table II best agreement with observation is found for a period of 3.2×10^4 sec. This applies to a star with $M = 30\, M_\odot$. In Table III

Table III. The Approach to Equilibrium—
Electron Capture *versus* Neutrino Loss Times (sec.)
in Pre-Supernovae Type II

$M/M_\odot$	50	30	21	15	10
ρ_6	1.1	3.1	6.3	12	28
t_e	9×10^4	3.2×10^4	1.6×10^4	8×10^3	4×10^3
t_ν	700	2×10^3	4×10^3	8×10^3	1.8×10^4

$$t_e \sim \rho^{-1} \sim \left(\frac{M}{M_\odot}\right)^2 \qquad Ni^{56} + 2e^- \rightarrow Fe^{56} + 2\nu$$

$$t_\nu \sim \rho \sim \left(\frac{M_\odot}{M}\right)^2 \qquad e^+ + e^- \rightarrow \nu + \bar{\nu}$$

electron capture times and neutrino loss times are given for stars of various masses. Exact correspondence is found for $M = 15\, M_\odot$. Type II supernovae are thought to result as the terminal explosive stage of stars in the mass range 10–50 $M_\odot$. The average *e*-process contribution from such stars properly weighted for the greater number stars in the lower end of the mass range will also show good correspondence with the observations.

Thus the relative abundances of the iron isotopes (and of the other *e*-process isotopes) strongly indicate a pre-supernova time scale for the *e*-process of the order of 10^3 to 10^4 sec. This is just what is to be expected if the $e^+ + e^- \rightarrow \nu + \bar{\nu}$ reaction occurs at the rate calculated on the assumption that it is governed by the universal interaction rate found for observable weak interactions including beta decay and muon decay and capture. This process has not been observed in the laboratory and the calculations are guided entirely by theory. The astrophysical evidence strongly implies that the theory is correct or alternatively that some unknown process leads to an energy loss comparable within a factor of 10 to that expected from electron-positron pair annihilation with neutrino-antineutrino emission.

The question is often asked: What becomes of the neutrinos and anti-neutrinos emitted by a star? The scientist may not know, but the poet does. Witness these lines:

> . . . . O dark dark dark. They all go into the dark,
> The vacant interstellar spaces, the vacant into the vacant,
>
> T. S. Eliot, *East Coker*

Beyond the iron-group nuclei, neutron capture processes have played the primary role in the synthesis in stars of the heavy elements. Because of repulsive Coulomb forces, charged particle reactions have been rather ineffective at the temperatures (10^8 to 10^9 degrees) at which the main line of heavy element synthesis has apparently occurred. The small relative abundance (0.1 to 1 per cent) of the lightest, "charge-rich" isotopes of the heavy elements attests to the infrequent operation of charged particle reactions in the synthesis of these elements. On the other hand neutrons interact rapidly with heavy nuclei at the "low" energies ($kT \backsim 10$ to 100 kev) corresponding to the temperatures just cited. In fact neutron reaction cross-sections vary roughly as $1/v \backsim 1/E^{\frac{1}{2}}$ where v is the neutron velocity and E the energy. Furthermore, at low energies the only reaction other than elastic scattering which is allowed energetically in most cases is the capture of the neutron. This leads to an increase in atomic weight by one unit, a slow but sure mechanism for the synthesis of heavier and heavier nuclei. Eventually, of course, neutron-induced fission becomes possible in the very heaviest nuclei at low energies. This process or alpha particle decay, or even spontaneous fission, depending on circumstances, terminates the synthesis.

Gamow, Alpher, and Herman suggested neutron capture as the mechanism of synthesis of all the elements starting with neutron decay in an early, highly condensed, high-temperature stage of the expanding universe. The density was taken to be $p \backsim 10^{-7}$ gm/cm^3 and the temperature to be $T \backsim 10^{10}$ degrees ($kT \backsim 1$ Mev). The measurements of Hughes and his collaborators on the capture of cross-sections of nuclei for fission spectrum neutrons in the Mev energy range indicated an inverse relationship between these cross-sections (σ) and isotopic abundances (N) such that $N \backsim 1/\sigma$. This was to be expected in general from the point of view of synthesis in a chain of successive neutron captures. Nuclei with small cross-sections would be expected to build up to large abundances in the chain, and vice versa, so that the number of captures per unit time would be uniform over contiguous sections of the chain. However, in recent years it has become clear from nuclear and astrophysical evidence that charged particle reactions must have played a considerable role in the synthesis of the *light* elements. For example, the iron-group abundance peak cannot be understood on the basis of neutron capture since the iron-group nuclei do not have anomalously low capture cross-sections.

Gamow's basic idea of neutron capture is incorporated in stellar synthesis, but the difficulties just mentioned are avoided by using charged particle reactions during various stages of stellar evolution to synthesize the elements up to and including the iron-group. Neutron production and capture then serve in the intermediate and terminal stages of stellar evolution as the main line of element synthesis beyond iron. In fact a small fraction, slightly over one tenth of 1 per cent, of the abundant iron-group nuclei are used as the "seed" nuclei at the start of the chain of captures. Mass spectroscopy has shown that the chain is unbroken in atomic mass in this region. (The chain is indeed unbroken beyond $A = 8$.) The abundance curve shows that two quite different and independent neutron capture processes have been necessary to synthesize the abundant isotopes of the heavy elements. In one of these processes, called the *s*-process, the neutron captures occur at a *slow* (*s*) rate compared to the intervening beta decays. Thus, the synthesis path lies along the bottom of the valley of mass stability and in general bypasses both the proton-rich, lightest isotopes and the neutron-rich, heaviest isotopes of the elements involved. On the other hand in the second neutron process, called the *r*-process, the neutron captures occur at a *rapid* (*r*) rate compared to beta decay. The captures lead rapidly from stable seed nuclei, predominantly Fe^{56}, to the very neutron-rich side of the mass valley and are stopped only by photoejection of the weakly bound neutrons by the ambient gamma-ray flux associated with the high temperature necessary for the production of the neutrons. Equilibrium between (n,γ) and (γ,n) reactions is established and progress along the synthesis path occurs only through electron-antineutrino ejection or beta decay which permits further neutron capture. On termination of the synthesizing neutron flux, the neutron-rich isobars at each atomic mass beta decay to the first stable isobar which then "shields" from *r*-process production those remaining isobars, if any, having fewer neutrons and more protons. The *s*-process and the *r*-process account in these ways for the synthesis of *all* the relatively abundant isotopes of the heavy elements. An exposure of a small fraction of the *s*- and *r*-process material to a hot proton flux or an intense photon flux will account for the production of the relatively rare, proton-rich, lighter isotopes of the heavy elements. This infrequent mechanism has been termed the *p*-process.

It follows from the evidence for two different neutron capture processes, which occur at quite different rates, that two separate and distinct stages of stellar evolution are demanded. The *s*-process has been assigned to the red giant stage of stars which were formed from galactic material containing light elements, particularly He, C, O, Ne, and Mg and the intermediate iron-group elements. These elements had been previously synthesized in other stars and ejected into the interstellar medium, mostly primordial hydrogen, of the Galaxy. The He, C, O, Ne, and Mg were required for the production of neutrons by α,n-reactions on C^{13}, O^{17}, Ne^{21}, Ne^{22}, Mg^{26} during the relatively slow helium burning in the red giant, with lifetimes 10^6 to 10^8 years. Professor Greenstein was the first to suggest a source of neutrons in stars—the exoergic $C^{13}(\alpha,n)O^{16}$ reaction.

The *r*-process is thought of as taking place in the exploding envelopes or cores of supernova outbursts. In this case the energy- and neutron-producing processes occur in the short interval of the supernova explosion, 1–100 sec., and the neutron captures accordingly occur at a rapid rate.

To illustrate the general importance of these considerations in determining isotope abundances, Figure 10 is appended. This shows the evidence for the operation of the three separate processes, *p, s,* and *r,* in the formation of the stable isotopes of the element tin. By following through the *s*-process path shown in Figure 11 and described in the figure caption, it will be seen that the first three isotopes, Sn^{112}, Sn^{114}, and Sn^{115}, cannot be made in the *s*- or the *r*-process. Their low abundances of the order of 1 per cent or less are consistent with their production only in the *p*-process.

Sn^{116} is the first isotope which can be made in the *s*-process, and the discontinuity in abundance between Sn^{115} and Sn^{116} is quite marked. Similarly, Sn^{120} is the last isotope which can be made in this process, and again, there is a discontinuity in going to Sn^{122} and Sn^{124} which can only be made in the *r*-process. The *r*-process apparently produced somewhat less abundances in this region of atomic weights than the *s*-process. This is a result of the "history"

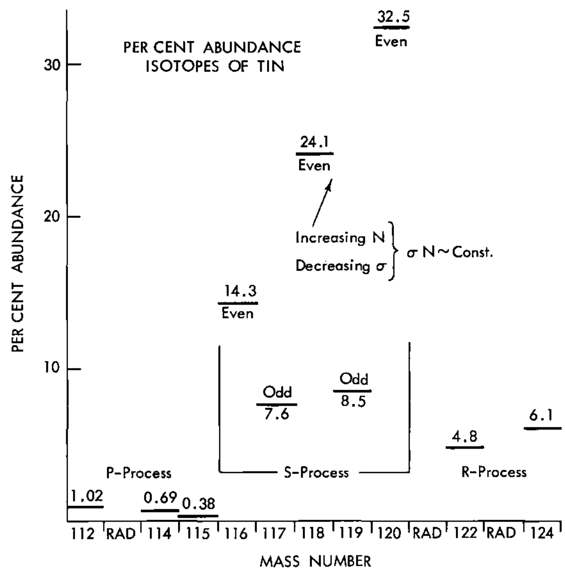


FIGURE 10. Abundance evidence for the operation of three separate processes, *p*, *s*, and *r*, in the formation of the stable isotopes of the element tin. The first three isotopes can only be produced in the relatively rare *p*-process involving charged particles (protons) or radiation, and their abundances are seen to be quite small. The next five isotopes are produced by neutron capture at a slow rate (*s*-process) and exhibit the regularity expected for this process-decreasing capture cross-section, hence increasing abundance, with increasing mass number. The last two isotopes are produced only by neutron capture at a rapid rate (*r*-process), and the discontinuity between the *s*-process and the *r*-process is quite apparent.

of the synthesis of the elements of the solar system, not of any fundamental nuclear properties of these isotopes. The rising trend in abundances from Sn^{116} to Sn^{120} is consistent with $N\sigma \backsim$ constant if we note that $\sigma(n,\gamma)$ in general decreases as more neutrons are added; and that σ, for odd A isotopes, is higher than σ for even A isotopes because of the tendency to pair up the neutrons. These statements are confirmed by recent measurements at Oak Ridge of 30 kev neutron capture cross-sections in separated tin isotopes. Measured cross-sections are compared with relative isotopic abundances for Sn^{116} to Sn^{124} in Table IV. The product σN varies over a factor of 2.6 for the first five isotopes and then drops by more than a factor of 10. This drop is explicable on the basis that Sn^{122} and Sn^{124} cannot be produced in the *s*-process and their σN should bear no particular relationship to that of the others. The variation in σN for Sn^{116} to Sn^{120} is puzzling until it is noted that Sn^{117} to Sn^{120} can be produced in the *r*-process as well as in the *s*-process, while Sn^{116} is produced *only* in the *s*-process. The *r*-production for Sn^{122} and Sn^{124} makes it possible to estimate N_r for Sn^{117} to Sn^{120} as given in the table. Then with $N_s = N - N_r$ it is possible to calculate σN_s which should be at most slowly varying over the range $116 \leq A \leq 120$ and indeed this is seen to be the case. There can be little doubt that the abundances

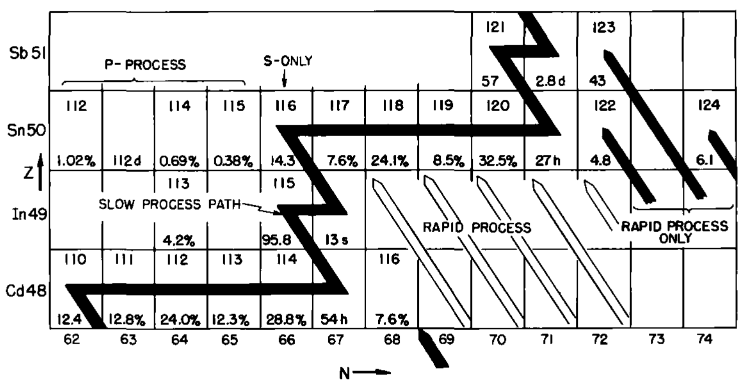


FIGURE 11. The *s*-process path through the isotopes of tin. The neutron number increases by units of one on a slow time scale until negative beta activity occurs and the path moves to the isobar of higher Z. This path can be determined from empirical evidence on the beta stability of nuclei. Note that the path bypasses the *p*-process and the *r*-process nuclei. The *r*-process nuclei are the end products of an isobaric beta decay chain, as shown at the far right from neutron-rich progenitors produced in an intense neutron flux. The *p*-process nuclei are produced by subjecting a small fraction of *s*- and *r*-process nuclei to an intense proton or photon flux.

and cross-sections of the tin isotopes reveal the nature of the processes by which they were synthesized. Isotopes of tin from red giants and supernovae once mixed together in the interstellar medium have not thereafter been separated and have come down to us as clues to stellar events in the distant past of the Galaxy. In other stars and other galaxies the relative *s*-process abundances will probably be much the same as in the solar system as will the relative *r*-process abundances but the ratio of $Sn^{122} + Sn^{124}$ to $Sn^{116} + Sn^{117} + Sn^{118} + Sn^{119} + Sn^{120}$ may show conspicuous variations relating to the past stellar history of the material involved.

The question often arises: Is there evidence that the *s*-process is occurring in present-day stars or that it has occurred recently? The most convincing answer is given in Figure 12 which shows spectroscopic evidence for the existence of technetium (Tc) in certain stars. Technetium has no stable isotopes and does not occur naturally on earth, but the isotope Tc^{99} is produced in the *s*-process chain and has a half-life of 2×10^5 years. If it had not been produced in the star and mixed to the surface in the last several hundred thousand years, it would have decayed to Ru^{99} and no technetium line would be observable. Tc^{97} and Tc^{98} have somewhat longer lifetimes but cannot be produced in the *s*-process. It is worthy of note that promethium has no long-lived isotopes and has not been observed in stars.

Turning attention for the moment to the *r*-process, detailed calculations reveal that it can account quantitatively for the abundance of the nuclei which

Table IV. σN FOR TIN ISOTOPES

Nucleus	process	σ (30 kev) (mb)	N(abundance) *per cent*	σN	N_r	$\sigma N_s = \sigma(N - N_r)$
Sn 116	*s*	92 ± 19	0.142	13.1	0	13.1
117	*sr*	390 ± 82	0.076	29.5	$\backsim$0.040	13.9
118	*sr*	59 ± 12	0.240	14.2	$\backsim$0.045	11.5
119	*sr*	243 ± 51	0.086	20.9	$\backsim$0.040	11.1
120	*sr*	35 ± 7	0.330	11.5	$\backsim$0.045	10.0
122	*r*	23 ± 5	0.047	1.1	0.047	–
124	*r*	23 ± 4	0.060	0.8	0.060	–

$$\frac{dN_A}{dt} = -\varphi_n\left(\sigma_A N_A - \sigma_{A-1} N_{A-1}\right) \approx 0 \qquad A \geq A(\text{seed nucleus})$$

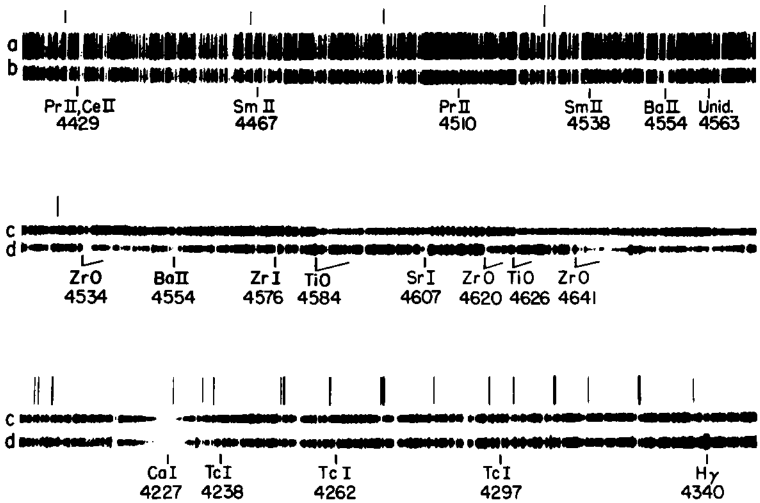


FIGURE 12. Portions of the spectra of stars showing the results of the *s*-process. *Top:* (*a*) Normal G type star, K Geminorum. (*b*) Ba II star, HD46407, showing the strengthening of the lines due to the *s*-process elements barium and some rare earths. *Middle:* (*c*) M type star, 56 Leonis, showing TiO bands at λλ 4584 and 4626. (*d*) S type star, R. Andromedae, showing ZrO bands which replace the TiO bands. Lines due to Sr I, Zr I, and Ba II are all strengthened. *Bottom:* (*c*) Another spectral region of the M type star, 56 Leonis; note that Tc I lines are weak or absent. (*d*) R. Andromedae; note the strong lines of Tc I.

can be made only or are most probably made in this way. Results of such calculations are shown in Figure 13. Combined calculations for the *r*-process and the *s*-process are shown in Figure 14. The correspondence between observations and calculations is in general fair although there are many unresolved problems pertaining to both.

It is possible to extend the calculations illustrated in Fig. 12 into the transuranic region and to calculate the abundance of Th^{232}, U^{235}, and U^{238} produced in each *r*-process event. These nuclei are singled out here because they are the parents of the naturally radioactive series with decay lifetimes comparable to astronomical times. This property has been used to arrive at determinations of the age of the meteorites, $\backsim 4.6 \times 10^9$ years, and of terrestrial rocks, $\lesssim 3 \times 10^9$ years. It is also possible to use these chronometers to measure the duration of stellar synthesis in the Galaxy once their production abundances are calculated. An important point is the fact that numerous short-lived progenitors of these nuclei are produced in

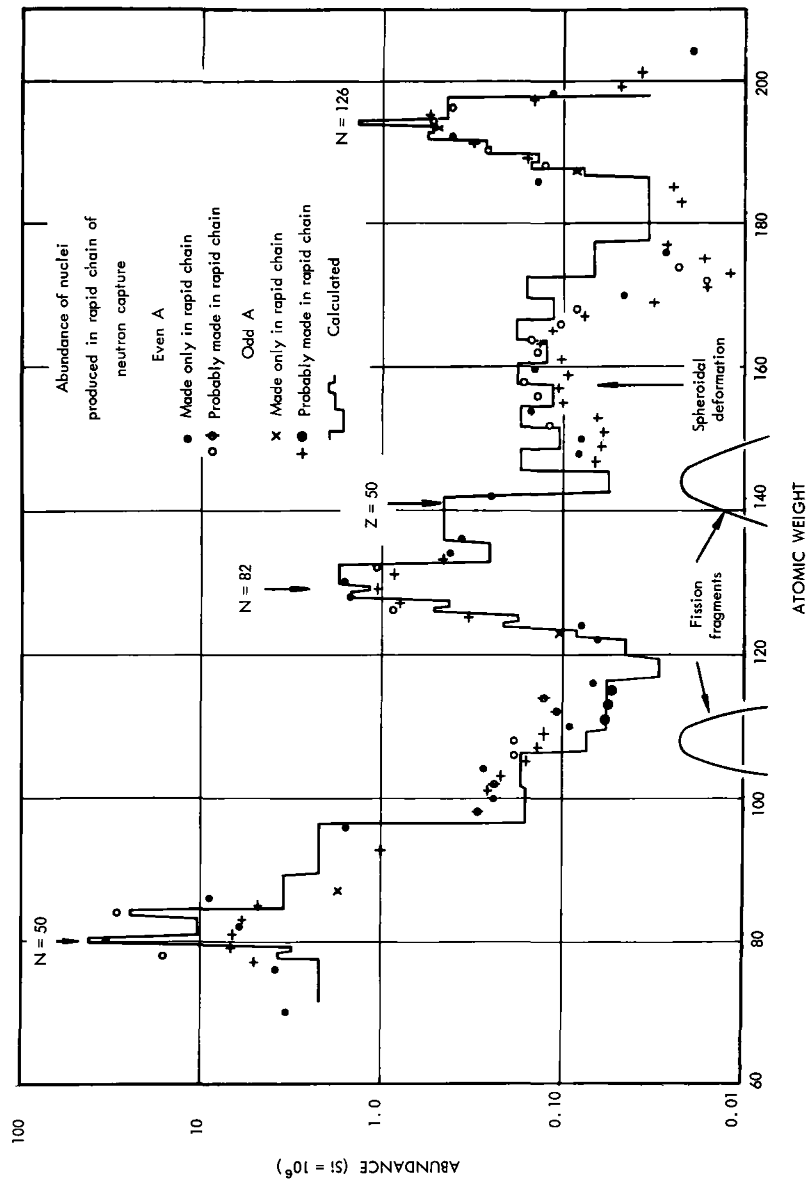


FIGURE 13. Abundances of nuclei produced in the *r*-process. The empirical points are taken from Suess and Urey. The histogram is a calculated curve. The free parameters have been adjusted to yield the correct relative heights of the three abundance peaks for magic neutron numbers $N = 50$, 82, and 126.

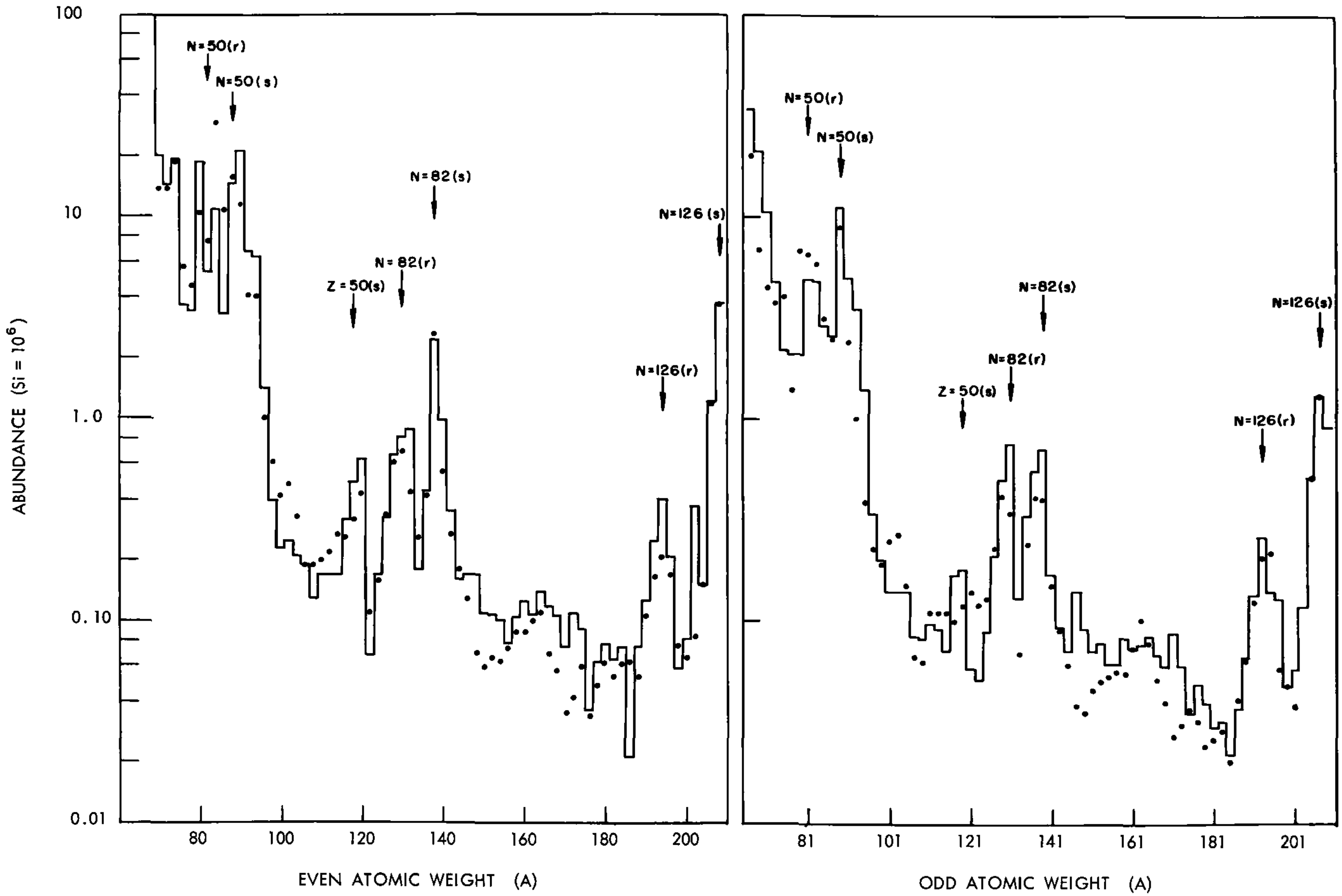


FIGURE 14. The observed meteoritic abundances (points) of the heavy nuclei compared with unsmoothed calculated values (histogram) based on the theory of the *r*-process and the *s*-process.

r-process events. Thus, for example, all the material produced at $A = 235$, 239, 243, 247, 251, and 255 contributed to the U^{235} abundance since the nuclei at $A = 239, 243, \ldots 255$ decay relatively rapidly by alpha particle and beta particle emission to U^{235}. At $A = 259$ and beyond, the ultimate fate of the nuclei is spontaneous fission rather than alpha decay and no contribution is made to the abundance of U^{235}. Contributions to U^{235} thus come from 6 progenitors. The situation is somewhat more complex for U^{238} and Th^{232} but the corresponding numbers are 3.1 and 5.75, respectively, as shown in Table V. Thus, on the basis of number of progenitors alone, the production ratio for U^{235} relative to U^{238} is 1.93 and that for Th^{232} to U^{238} is 1.85. Detailed calculations yield 1.65 ± 0.15 for each of these ratios, the exact agreement being accidental. Figure 15 shows the use to which these ratios can be put. The present-day observed ratios in meteorites are $Th^{232}/U^{238} = 3.8$ and $U^{235}/U^{238} = 0.0072$, as indicated in the figure. The *differential* mean lifetimes are 9.63×10^9 years for Th^{232} *versus* U^{238}, and 1.22×10^9 years for U^{235} *versus* U^{238}. The abundance ratios extended back in time to the origin of the solar system are simply straight lines on the logarithmic abundance scale of Figure 15. For a single sudden synthesis event at some time in the remote past, the extensions to that time are also straight lines back to the relative production ratio 1.65 ± 0.15. The dates so determined are discordant by $\backsim 2 + 10^9$

Table V. PROGENITORS OF U AND TH

Parent	U^{238}	U^{235}	Th^{232}
Great, ... granddaughter	Pb^{206}	Pb^{207}	Pb^{208}
Mean lifetime (10^9 yr)	6.51	1.03	20.1
Progenitors (A)	238	235	232
	242 (α)*	239 (α)	236 (α)
	246 (α)	243 (α)	240 (α)
	250 (10 per cent α)	247 (α)	244 (α)
		251 (α)	248 (89 per cent α)
		255 (α)	252 (97 per cent α)‡
	254 (SF)	259 (SF)	256 (SF)
Total number	3.1	6	5.75
Ratio of progenitors	1	1.93	1.85
Calculated abundance ratio	1	1.65 ± 0.15	1.65 ± 0.15

* In this table α designates alpha particle decay, while SF designates decay by spontaneous fission.
‡ The yield at $A = 252$ must be multiplied both by 0.97 and by 0.89 to give the fraction which ultimately becomes Th^{232}.

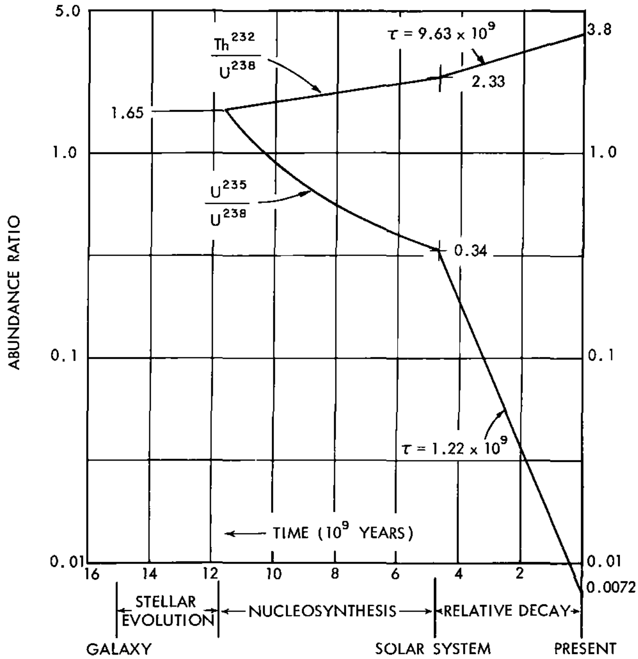


FIGURE 15. The abundance ratios Th^{232}/U^{238} and U^{235}/U^{238} as a function of time for solar system material before and after formation of the system. The curves are adjusted to the present-day ratio and to the calculated ratio for production in each *r*-process event. The calculations indicate continuous synthesis in the Galaxy from 12×10^9 to 4.7×10^9 years ago. Type I supernovae are assigned an evolution lifetime of 3×10^9 years leading to an age for the Galaxy of 15×10^9 years. If the *r*-process occurs in rapidly evolving massive condensations, then the age is 12×10^9 years.

years. The curves shown are those expected for uniform synthesis over the nucleosynthesis interval in the Galaxy. Concordant results are obtained from the Th^{232}/U^{238} and the U^{235}/U^{238} ratios at a time for the beginning of *r*-process nucleosynthesis some 12×10^9 years ago. This is a lower limit for the age of the Galaxy since it may be necessary to add an interval of as much as 3×10^9 years for the time for stars to evolve to the supernovae in which the *r*-process occurs. On the other hand, the *r*-process may have occurred in violent explosions of massive condensations which occurred at the formation of the Galaxy. In spite of this and other uncertainties, the abundances of Th^{232}, U^{235}, and U^{238} found on the earth point to an age of the Galaxy somewhere in the interval from 10 to 15×10^9 years. One must not overlook, however, the comment of Samuel

Pepys when the date of Genesis calculated from references in Scripture came under question by the geologists of his time. Said Pepys:

> To the Rhenish wine house, and there came Jonas Moore, the mathematician, to us, ... and spoke very many things, not so much to prove the Scripture false as that the time therein is not well computed nor understood.
>
> *The Diary of Samuel Pepys,* May 23, 1661

In conclusion, it is necessary to mention a new development in astrophysics which almost certainly has important ramifications in regard to nucleosynthesis in the Galaxy. Large radio sources associated with certain galaxies are found to radiate at rates approaching 10^{44} ergs sec. $^{-1}$, some 10^{11} times the optical luminosity of the sun. If this radiation is due to the synchrotron mechanism, the energy stored in the magnetic field and the high-energy electrons circulating in the field lies in the range 10^{60} to 10^{62} ergs corresponding to the rest mass energy of 10^6 to 10^8 solar rest masses. Hoyle and Fowler have suggested that this energy was made available at the expense of gravitational energy during rapid collapse of objects with masses somewhat greater than the range just indicated. The red shift measured in the optical emission by the so-called radio stars places them at a great distance from the Galaxy and implies total optical luminosities up to 10^{46} ergs sec.$^{-1}$ for these objects. Observations on M–82 by Sandage and Lynds show that an explosion involving 5×10^6 solar masses is occurring in that galaxy. The Burbidges have found evidence for "violent events" in numerous galaxies. It can be expected that nuclear reactions will take place during such events. The production of large amounts of helium and lesser amounts of heavier elements may well occur in this way during the formative stages or early history of galaxies. In this way it may be possible to understand the apparent universality of the helium-to-hydrogen ratio, $\backsim 10$ per cent by number, in the stars of the Galaxy independent of age and position. Massive stars produced helium early in the history of the Galaxy; less massive stars with longer evolution times produced the bulk of the elements heavier than helium. In a way these violent galactic events play a role in the Hoyle-Bondi-Gold cosmology similar to the universal "big bang" in Gamow's cosmology. It may not be too trite to say that the Universe is all things to all men!

✲

The History of Stars and Galaxies

JESSE L. GREENSTEIN

THE UNIVERSE is enormous, strange, and untouchable; man's technical means and intellect are small and short-lived. Discussing stellar or galactic evolution is a large task made no easier by the lack of astronomical meaning in such commonly used words as history, evolution, birth, life, and death of atoms and stars. Let us maintain belief only in presently known sources of energy and in the irreversibility of the second law of thermodynamics. We have no definite observational evidence that the expansion of the universe will reverse itself, or that matter and energy now appear out of the vacuum. Our locally observable universe is on a one-way road 10 billion years in length, the same for the oldest atomic nuclei on the earth and for the oldest groups of stars in our Galaxy as that indicated by the redshift of distant galaxies. This current agreement may be an optimistic accident, since the certainty of many of our age-dating techniques is not high. Also, we have changed our minds very often. Nor is there as yet any positive, generally accepted disproof of the steady-state hypothesis.

But such an age brings into clear relief the modifications required in the ideas of organic evolution. The solar nuclear energy sources are sufficient to last for a total of 10 billion years, and the sun is about 5 billion years old. Massive, high luminosity stars live only a million years. Less massive stars are much less luminous and have lives of trillions of years. Stars exist which may be nearly as old as the atoms they contain, yet other stars must have been recently born. For the faint stars, evolution does not exist, history has only begun, the species and the individual have the same life span. For the bright stars to exist, there must be a reservoir of matter (and nuclear fuel) out of which they continue to be born—the interstellar gas. If it is not to be rapidly exhausted, and if the general picture given by the current nucleosynthesis theory is correct, this gas must itself be replenished. We believe that matter initially consisted largely of hydrogen or neutrons (and

JESSE L. GREENSTEIN Mount Wilson and Palomar Observatories, California Institute of Technology

possibly some helium if there was initial synthesis in the early stages of the expansion of the universe), that there has been a secular, probably non-uniform, increase in the helium and metal content of the interstellar gas, and that the stars formed later in the history of the Galaxy. Consequently, we require a cyclic flow from interstellar gas, via star formation, to the hot interiors of the stars, and a significant mass-loss by the stars back to the gas phase. Exhaustion of nuclear fuel (largely hydrogen) and the slow, small change in the composition of the gas is the only real history of stars; as for the universe in the large, its history is its expansion. Stellar evolution is concerned with the behavior of a single star during its life span; *i.e.*, the effects on the star's luminosity, radius, and surface temperature of the change in its chemical composition and possibly in its mass.

This subject has a short time scale, and we are in a period of rapid growth of sophistication and insight. The first mathematical treatment of a star considered to be a self-gravitating sphere of gas was published in 1870 by J. Homer Lane, an isolated pioneer. Lane, a lone figure, became a member of our Academy. He obtained the first reasonable estimates of the temperature of the sun and its internal structure, and also developed similarity transformation rules that make it easy to obtain a general insight into the behavior of a wide variety of stars. As far as energy sources were concerned, Lord Kelvin in 1861, following Helmholtz in 1854, evaluated the thermal and radiated energy derived from gravitational energy during the slow collapse of a gas sphere. He found it to be sufficient to maintain the solar luminosity for only 10^7 years. Fifty years ago, before nuclear physics really existed, Harkins and Wilson (1915), and Eddington in detail (1920), discussed the astronomical evidence for "subatomic" energy, a subject brought to the modern phase only 25 years ago by Bethe (1938). Thus, while the study of stellar interiors and the Academy have nearly the same age, the rapid advances in our subject, based on sound physics, are quite recent.

The astronomical universe contains many surprises, and there are two possible approaches to its problems. One is to discuss the normal processes of stellar and galaxy evolution, applying that part of physics we now think relevant. We must first neglect the important non-thermal aspects (magnetic fields, high-energy particle acceleration, cosmic rays, and giant stellar and superstellar explosions). We cannot ask questions concerning the origin of the universe and its expansion and the possible unknown links between geometry, gravitation, and nuclear physics which may be involved in the birth of new matter and energy. I will therefore discuss first the normal processes, but will return, finally, to some difficult and speculative questions for the Academy to consider during its next hundred years.

THE STARS

Consider the equilibrium structure of a star where gravity is balanced by gas pressure. Counting the number of boundary conditions and equations, we find that a stellar interior structure, i.e. density, temperature as a function of radial distance from the center, is completely specified by the mass of the star and its chemical composition. For the latter we need only the abundance of hydrogen and helium, and of C, N, O, Ne and heavy elements as a group, except that we must allow for the possible difference of composition from surface to core. This variation, largely in the H/He ratio, or H/He/CNONe ratios, can be caused by the nuclear transformation of hydrogen to helium, or at later stages of He into C, O, Ne. If there is too little driving force to produce convective mixing from core to surface, the inhomogeneity of the chemical composition causes drastic alterations in the radius, R, and luminosity, L, of the star, which are the observable parameters. (The surface temperature, T, is derived from R and L.) Stars are located uniquely, once mass and composition are given, in an L, T diagram, or in various transforms of this, as for example, the so-called Hertzsprung-Russell diagram. Obviously, additional effects caused by convection, composition, mixing, rotation, or magnetic fields will occur; some of them are predicted by the mass and composition, others may be complex results of the environment out of which the star was born. We will trace the current picture of normal evolution and then later of catastrophes which some stars may encounter.

Stars are formed from denser than average gas and dust clouds. The density fluctuation must be appreciable if the protostar is to be stable against a disruptive environment of tidal disturbances and shear caused by galactic rotation. As a star contracts, it finds itself with excessive angular momentum and magnetic field stresses. Some types of stars rotate rapidly, others like the sun, only very slowly. We now believe that the process of disposal of excessive angular momentum and magnetic field is connected with the origin of planetary systems. The evolution of a star from a density fluctuation in the gas is a complex process dependent partly on the opacity of the gas and dust at low temperatures and on chemistry as well.

We have evidence that stars exist and tend to be formed in groups or clusters of from 100 to 100,000 members—just as galaxies also tend to occur in groups. The formation process may be connected with the birth of massive stars of high luminosity. Their nuclear energy, converted in part into ultraviolet radiation, can produce an ionization and pressure wave or shock wave, providing the needed density increases in a large volume of the gas cloud. Stars are formed with a variety of masses. The probability distribution of the masses formed can be derived from the observed distribution function of stellar masses after correction for the short life span of the massive stars. At present, only 10^{-5} of the stars have a mass 25 times greater than the sun, but probably one star in 10^2 is formed with such a mass. However, there is an upper limit to the mass of normal, stable stars, because the radiation pressure exceeds gas pressure and a radiation gas is only marginally stable. This limit is now thought to be near 65 solar masses. What happens to larger masses will be of great importance in the later discussion of superexplosions.

Once a protostar begins to contract, its central temperature rises because of conversion of part of the gravitational potential energy into internal energy, the balance being radiated. Nuclear reactions do not occur until nearly 10^6 °K, so that the early history depends on the gravitational contraction. Figure 1 shows recent computations of evolutionary tracks during early gravitational collapse; these results by Hayashi differ greatly from those based on a Helmholtz-Kelvin time scale derived at constant luminosity. On this most recent hypothesis stars, soon after the earliest stages of formation, suffer violent convection which causes them to have a much higher luminosity than when they reach their equilibrium configuration on the main sequence. The loci of equal ages traverse regions of the L, T diagram where young clusters now contain unstable variable stars, which may now also be in the throes of planet formation. We have observed that they contain large amounts of lithium. We believe that this is evidence of high-energy nuclear spallation connected with the release of magnetic and rotational energy during star and planet birth.

Once a star is stabilized in its equilibrium configuration, its life on the main sequence is set by the luminosity to mass ratio, which is a steep function of mass. The total lifetime for nuclear sources is nearly 10^{10} M/L years (where M and L are in solar units). For the faintest red dwarfs this is about 5×10^{12} years; such stars once formed do not evolve on any reasonable time scale and remove matter from

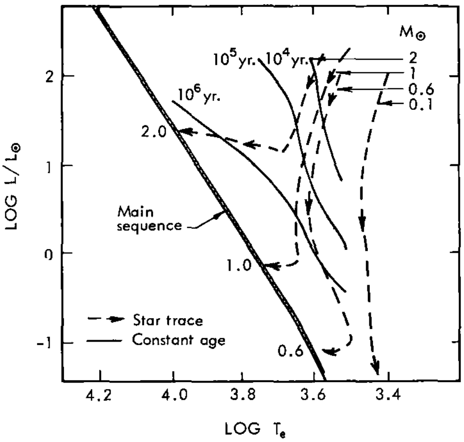


FIGURE 1. Schematic, nearly vertical, evolutionary tracks (dashed) of stars of different masses, approaching the main sequence. The loci of constant age are given, and lie in the region where stars in young galactic clusters are found.

the re-cycling process between interstellar gas and stars. The least massive do not even produce nuclear energy and may become degenerate dwarfs without consuming their nuclear fuel. The faint red dwarfs are the most common objects in the Galaxy and will, in the future, form an even larger percentage. At present with star birth and death in near balance, our Galaxy should be decreasing 8 per cent in brightness per billion years. But the faint red dwarfs, less massive than 0.3 ⊙, which will still be shining in 100 billion years, now contribute only 0.4 per cent of the light. So our Milky Way and our neighboring galaxies fade into uninteresting ghosts.

But not without fireworks. The spectacular stars of large mass provide most of the light now, and will continue to do so until the nuclear fuel in the interstellar gas is exhausted. They evolve very rapidly and are the subject of much current study and large machine-computing programs. The evolutionary tracks start away from the main sequence in 10^{10} M/L years (which may be as little as 10^5 or 10^6 years) and depend in particular on the inhomogeneity of molecular weight caused by conversion of hydrogen into helium in the core. For less massive stars, the metal to hydrogen ratio is important, so that the evolution of old, metal-poor stars in globular clusters differs from that of the younger stars in our neighborhood. Such differences appear in the L, T diagrams of the various population types. The central temperatures of such stars are high enough, after hydrogen is consumed, for nuclear reactions between alpha particles and even

FIGURE 2. A young galactic cluster is immersed in a gas cloud (NGC 2264). High luminosity, hot stars cause the hydrogen to fluoresce; the region also still contains faint stars contracting toward the main sequence.

carbon nuclei to occur in the red giants and supergiants. Under such conditions, the question is how inhomogeneous the interior can remain as successive energy sources are tapped. The rate of convective mixing driven by an oversteep temperature gradient remains as a major uncertainty in stellar hydrodynamics, and in the final evolutionary tracks. The current view is given schematically in Figure 3 for massive stars of the young population. Old stars of low metal content with masses larger than 1.5 $M\odot$ no longer exist. Their evolution starts fainter and proceeds more nearly vertically; since stars now evolving in globular clusters have a mass near that of the sun, time scales of over 10 billion years are indicated. The difference of age between various young groups of galactic clusters produces the observed *L, T* diagrams shown in Figure 4. For globular clusters, all of which are old, the main variable is chemical composition. Semi-empirical evolutionary tracks for globular clusters are shown in Figure 5. Figure 6 gives the *L, T* diagram for nearby stars in our own Galaxy; it shows that the evolved old stars resemble the globular clusters and the very old galactic cluster population. (Because of selection effects, this figure does not give a true picture of the relative frequencies of different types of stars.)

During the complicated evolution of massive stars, the central temperature goes through certain critical stages for nuclear energy production and element synthesis. A star near solar mass has a central temperature from 10^7 to 10^8 °K; at 4 $M\odot$ the temperatures are three times higher, while at 16 $M\odot$, they start at

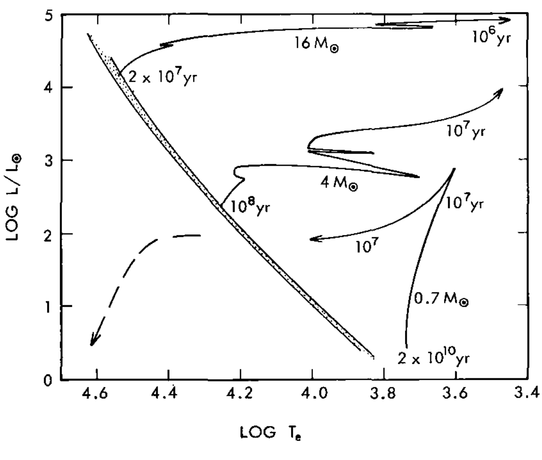


FIGURE 3. The theoretical evolution of massive stars, once hydrogen is exhausted, is nearly horizontal and towards the right. Ultimately, such stars meet catastrophic ends, or return to the lower left. The tracks are uncertain in detail, but pass through regions of the *L,T* diagram where supergiants are found. The track for 0.7 $M\odot$ runs upwards at first, and then to the left.

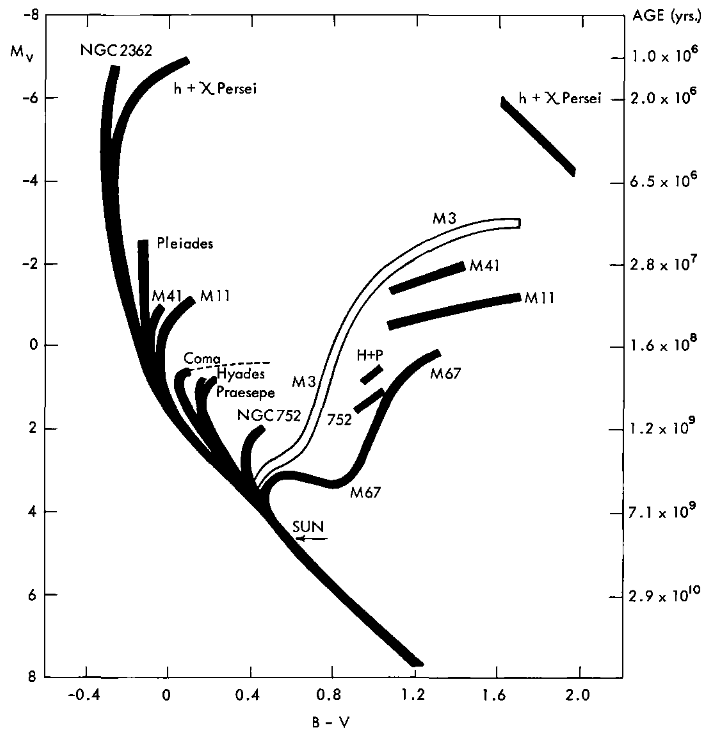


FIGURE 4. Young galactic clusters of stars occupy different regions of the L, T diagram dependent on age. The most luminous stars exhaust nuclear energy early, and deviate from the main sequence toward the right. An age scale (subject to modification) is given at the right. On the left, M_v is the absolute magnitude; B-V is a measure of color, with hot blue stars on the left.

3×10^7 and approach 10^9 °K. I have recently tried to find evidence in the surface composition of the stars for effects of these various temperature regimes. Since there is little mixing between core and surface, the composition changes occur only in certain stages. Stars with substantial mass loss form such a group, as do some red giants. The white dwarfs have masses less than 1.2 $M_\odot$, and some are apparently the exposed cores of red giants. About 10 per cent of the white dwarfs have no hydrogen left, and have atmospheres of helium and carbon. The red-giant variables contain two major peculiar groups. The first has a very large increase in carbon and probably in helium abundances. The second has excess abundance of the elements of atomic number 38 to 72, especially of Zr and the

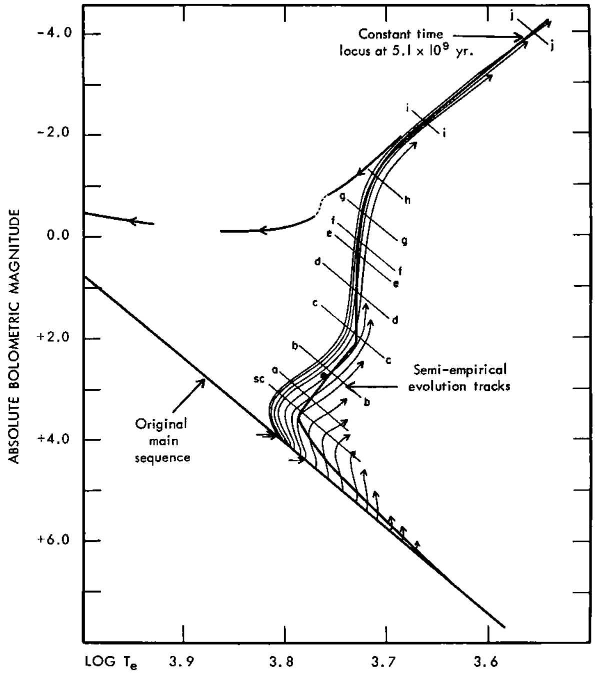


FIGURE 5. The globular clusters are very old. Stars slightly more massive than the sun have already begun to evolve upward and to the right. The arrows show how evolving main-sequence stars of low metal content are mapped onto the globular cluster L, T diagram.

rare earths. This group also contains Tc^{43}, an unstable element. For helium and carbon to replace hydrogen requires temperatures of 10^7 and 3×10^8 °K, respectively; such direct observations give us, I hope, some support for our ideas of stellar structure and perhaps even some confidence in nuclear physics. The formation of the heavier elements from the abundant iron-group nuclei is by neutron capture. The synthesis of C^{12} from He^4 is the key, since stable isotopes with an easily removable neutron can be formed by exposure of C^{12}, O^{16}, Ne^{20} to protons. Then C^{13} or Ne^{21} plus alpha particles gives the neutrons at only 10^8 °K. At higher temperatures, as Cameron, Fowler, Hoyle, and the Burbidges have shown, heavier elements are synthesized.

However, above 5×10^8 °K, stellar evolution usually becomes catastrophic. Energy sources are meager and the star heats by gravitational collapse, implodes, and then explodes. Neutrinos, which cannot be absorbed by the star, are produced by beta decays and by photo-neutrino and pair-annihilation processes. Neutrinos carry away only a negligible fraction of the energy in normal stars, but become a major loss for stars of large mass. Fowler and others have discussed the disasters that massive stars encounter — collapse and supernova explosions. The spectacular outbursts in which a star emits in a month as much light as the sun does in 10 billion years are being photometrically and spectroscopically studied. No possible stellar supernova remnant has been certainly identified. The expanding gas clouds, however, are associated with radio sources in our Galaxy like the Crab nebula, Cas A and the Cygnus loop, and inject the reaction products of heavy element nucleosynthesis into the interstellar gas. They are possibly also the injection mechanism for the cosmic rays. They stir the interstellar gas and may

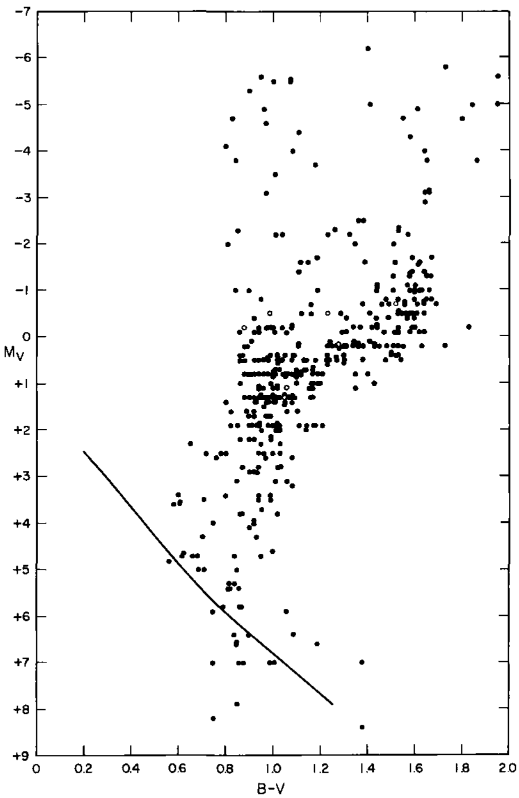


FIGURE 6. Many of the brighter stars near the sun have an L,T diagram (from O. C. Wilson) which shows that they have evolved to the right and above the main sequence. They resemble stars in globular clusters, or in the oldest galactic clusters, but are sometimes more luminous, (*i.e.* more massive). The faintest, and unevolved stars lie along the main sequence (solid line).

even trigger star formation in galaxies rich in gas. The total energy density of neutrinos produced by supernovae at their present rate of occurrence is small, although it remains a speculative possibility that large numbers of neutrinos, because of gravitational collapse of massive stars, were formed early in the history of a galaxy and may still be an important contribution to the total matter-energy density of the universe.

In the precollapse stage, the massive stars evolve rapidly. Stars as luminous as 10^6 $L\odot$ are found in clusters in our own Galaxy and in gas-rich spiral nebulae. The number of high luminosity stars present depends on the balance between the rate of star formation and their lifetime. In galactic clusters, once star formation has ceased, the brightest stars burn off, traveling across the L, T diagram into the red giant region of fainter and fainter stars as the cluster ages. The globular clusters are all nearly the same age, and stars of about solar brightness are reaching the end of their main-sequence life span and becoming red giants. The major characteristics of the old stellar populations derive from the early cessation of star formation in the gas-poor old population at a time when the metal content was low.

The final stages of stellar evolution, when nuclear fuel is gone, involve a considerable loss of mass from the red giant. A rapid transition, as yet unobserved, into the hot, small-star region follows. Gravitational collapse is stopped by the degenerate-electron gas pressure. The star must be less massive than 1.2 $\odot$ (if made of helium or heavier elements) if the stable configuration of the white dwarfs is to be reached. The white dwarfs radiate only part of the thermal energy of the non-degenerate nuclei, but this is sufficient for a very long time scale. At constant radius the surface temperature drops, the luminosity decreases, and the duration of a given brightness-stage increases. There are only three white dwarfs known, of a luminosity less than 10^{-4} $L\odot$ and a surface temperature below 6000°K. To fade into the invisibility of infrared emission only, temperatures of 1500°K or less are required. But the time to reach such a "black-dwarf" stage is 200 billion years. There may be a direct evolutionary path for gas condensations of mass less than 0.1 $M\odot$ to form hydrogen-rich degenerate objects, too cool for nuclear energy production. But we do not expect the Galaxy to contain much mass in such old, invisible stars, and it is still too young for the white dwarfs to dominate, as they eventually will. The sun as a whole emits 2 ergs/gm sec.; certain galaxies have 50 times lower emission per gram.

We do not know whether a considerable amount of gravitational mass exists in some galaxies, which cannot be accounted for. There is still a speculative possibility that old, very massive stars have undergone gravitational collapse into superdense, invisible stars. It is a mundane, but more probable alternative that certain systems contain large numbers of stars of small mass and low luminosity.

THE EVOLUTION OF GALAXIES

Two hundred years ago the philosophers were much concerned with the conception of island universes; Thomas Wright of Durham and Immanuel Kant suggested the modern concept and Herschel's observations provided numerous examples of possible systems of stars outside the Milky Way. Herbert Spencer popularized the idea exactly 100 years ago. But it is only 40 years since individual stars were recognized in other galaxies and the large velocities of recession observed. I will not discuss the evolution of the universe as a whole, but will limit myself to that of individual galaxies. Obviously, we must make one assumption of great uncertainty, that most of the matter-energy in the universe is located inside galaxies. We can photograph about 10^8 of these stellar aggregates, out to 5 billion light years, but they fill only a millionth of this volume. Is there much intergalactic material? Are galaxies essentially closed systems, or do they exchange gas and stars with a low density environment? Are new galaxies still being formed from invisible, as yet undetectable hydrogen? Such unanswered questions are part of the work of the next hundred years.

The amount of detail we use in describing a system of stars decreases enormously with its distance. The nearest members of the local group of galaxies are essentially resolvable into stars, but only the intrinsically brightest stars are visible. In the crowded central regions of our near twin, the Andromeda nebula, the stars are too crowded to be resolved. In no galaxy can stars as faint as the sun be studied. We may observe the bright, rapidly evolving stars in nearby systems, but only the global properties of distant systems. Given the initial formation rate of stars and the rate of exhaustion of the interstellar gas, we should in principle be able to predict what the L, T diagram of a system looks like as a function of age. We are far from being able to do this and have few observational facts.

The nuclear history of an assemblage of stars is only one part of the history of galaxies. In them we see, in the large, the distribution of total luminosity across the object in radius and angle, since most galaxies are not symmetric. The stars move in their mutual gravitational potential field, largely without collision. The structure of a galaxy depends on the balance between gravity and its angular momentum (which may be large or apparently nearly zero). The gas density and possibly the magnetic field govern the appearance of spiral arms, if any exist. The great variety of galaxy types observed has been fairly well systematized, and some reasonable evolutionary patterns deduced, based on a combination of morphological characteristics and current ideas of stellar evolution.

A major difference in the stellar evolution pattern distinguishes the nearly spherical from the highly flattened galaxies. Members of the first group, called ellipticals, appear to have consumed most of their original gas content quickly, so that star formation lasted only a short time. The system of stars, with a wide range in masses, has had a relatively simple history of burning-out stars of successively lower luminosities. The brightest stars are red giants of what is called Population II, similar to but not identical with those in globular clusters. The better but not complete analogy is with very old galactic clusters, which also do not show the great metal deficiency of some globular clusters. The ellipticals are nearly dust-free but have some interstellar gas, which may come from stellar mass loss, or from the intergalactic medium. They have a low average value of luminosity per gram, less than one-tenth that of our Galaxy, arising in part from the absence of new, high-luminosity stars, but possibly also from a larger frequency of stars of low mass than in the solar neighborhood. Since the giant ellipticals are the most massive and luminous galaxies known, they are important for the red-shift-distance relation and cosmological problems. The evolutionary change of their brightness, as we look back in time, can only be estimated; they must have been brighter, perhaps even by a factor of two, near the limit of our present surveys. This quite uncertain correction is of the order of the differences between various relativistic models, or between them and the steady-state theory. It must be known better if important cosmological problems are to be seriously attacked. Were we able to look far enough back in time, say nine-tenths the "age of the universe," we would see the ellipticals in their earliest stages of evolution, where they should have been very much brighter. Clearly, one certain disproof of the steady-state theory would be the discovery of such an evolution backward in time. A significant difference between the average properties of the universe now (as we observed it in our neighborhood) and a long while ago (at great distances) should exist in "explosive" or evolutionary cosmologies. Such a decision, however, must depend on very obscure details of interstellar gas density, star formation, and evolution. Other tests such as the frequency of distant radio galaxies involve even more uncertain hypotheses.

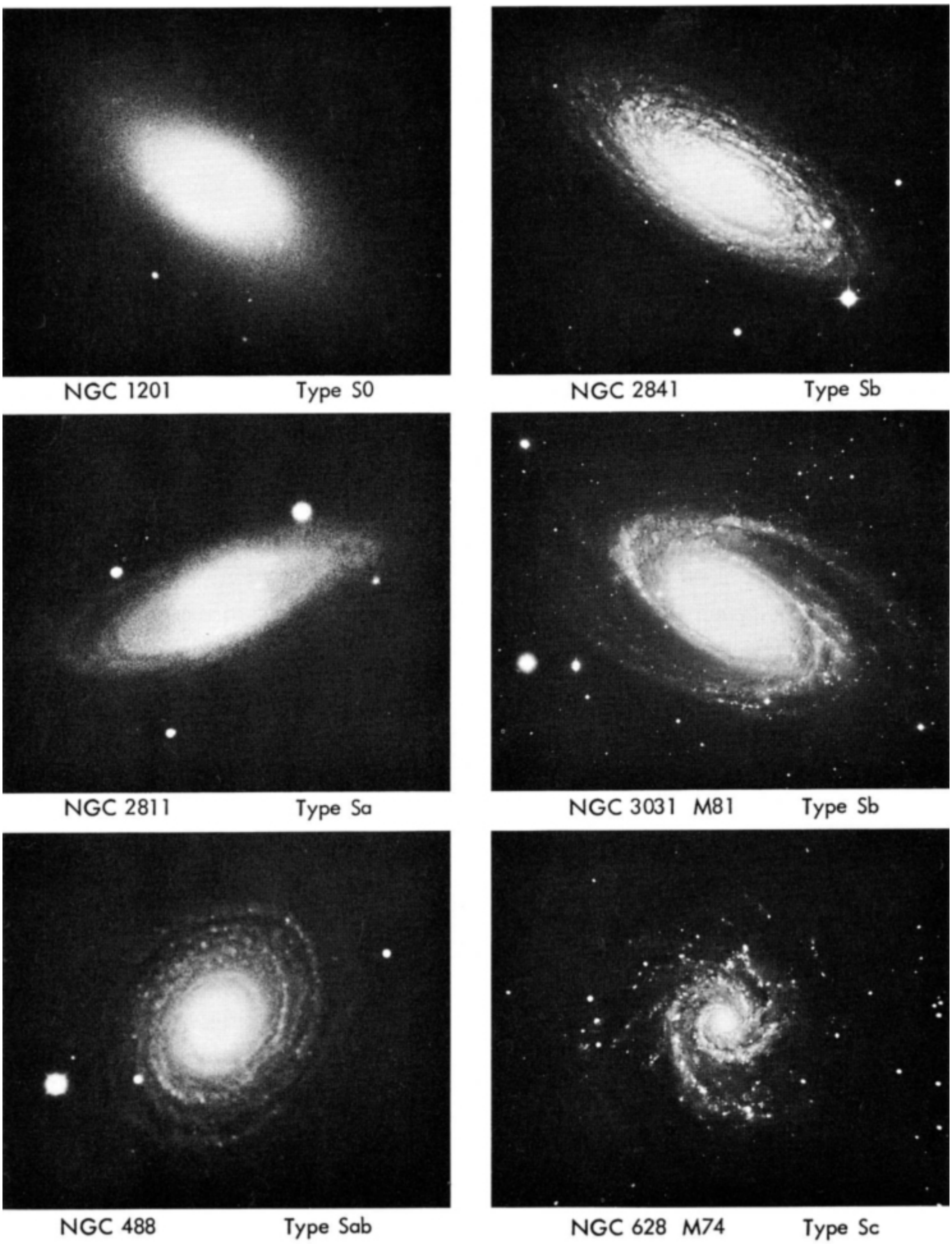
The spiral nebulae which contain varied amounts of the young Population I and the old Population II differ systematically from the ellipticals in the larger angular momentum of part of their structure, the characteristic thin galactic disk. The various types of elliptical nebulae have ratios of minor to major axes greater than 0.3. Only gas-rich spirals are more flattened. Not all components of a spiral have equal momentum — a spiral, in fact, consists of a set of subsystems of varied degrees of flattening, immersed in a nearly spherical halo. The least flattened subsystem contains the oldest objects, with low metal content, such as the globular clusters and very high-velocity stars. In this group, star formation lasted only a short while, finishing soon after it began. Present theories of element building and our observations of stellar composition of the oldest galactic-halo stars suggest that, while helium and heavy-metal synthesis began early and rapidly, it was only partially completed when the oldest stars were formed. No object with less than one two-hundredth of the metal content of the sun is known, and no very great helium deficiencies are observed. Gas with larger angular momentum collapsed rapidly towards the equatorial plane, because of viscous energy losses. In our Galaxy, the residual gas is largely within a cylinder whose thickness is only 1 per cent of its diameter. Star formation has continued for 10 billion years without large change in composition and it is within this disk that the spiral arms, gas, and dust are found. Clearly, a delicate balance is necessary if any gas is to remain. Too rapid star formation would have locked most of the gas in low-luminosity, non-evolving stars. It is likely that star formation occurs in bursts, and that the spiral structure is short-lived. The gas is heated by high-luminosity hot stars; the magnetic field and the shear produced by galactic rotation also resist gravitational compression and prevent growth of density fluctuations into stars.

Unless stars and gas with high angular momentum can be lost into space, there is no obvious evolution which can transform a spiral into an elliptical galaxy. I am indebted to Dr. Allan Sandage for some of the following general concepts illustrated by his remarkable collection of photographs in the *Hubble Atlas of Galaxies*. We have begun to understand the different types of spirals; the probable direction of evolution is from the nearly amorphous irregular galaxies to the very open spirals, through progressive stages where the arms become thinner, more nearly circular, and tightly wound. The spiral arms are widely spaced initially and have a very steep pitch-angle and only a small nucleus of star in the so-called *Sc* types. The highest-luminosity young stars are found in such objects. The rotation period of a galaxy is about 2×10^8 years, and it is interesting to note that few spiral arms can be traced for more than two complete turns; i.e., if they were simply winding up and merging into the background they would last only 400 million years (see Figure 7). Arms must either re-form out of the gas, be a locus of disturbance in a substratum, or flow out of the central regions, somehow acquiring the angular momentum. We directly observe only neutral or ionized hydrogen in our Galaxy; we do not know whether there is a large mass of molecular hydrogen to serve as a reservoir. There is little atomic hydrogen or dust visible between spiral arms. Molecular hydrogen forms only with difficulty at low space densities. It should therefore also be concentrated within spiral arms, possibly only on the surface of dust grains. Of course, we do not know the concentration of hydrogen molecules in intergalactic space, leaving still another unknown in the cosmological problem.

Returning to evolution in spirals, the *Sc* and *Sb* spiral nebulae have progressively larger central regions containing many of the old disk population stars (like those found in elliptical nebulae). The arms are less clearly marked, have lower contrast, and the luminosity of the brightest stars drops. The *Sa* spirals are so tightly wound that they resemble a disk with a central bright core. They merge into the *SO*, which have no gas or bright stars, but still show faintly the dust arms left. The most flattened ellipticals, the *E7* group, are close to the *SO*. Thus the morphology shown in Figure 8 suggests an evolution of rapidly rotating systems, forming and re-forming spiral arms, building up a larger population of evolved red giants, and ending in the *SO* type. No direct connection into the elliptical nebulae, however, is known. The ellipticals are much more massive and have low light output per unit mass. We cannot guarantee that a galaxy is a closed system during this evolution.

Radio observations at the 21-cm wavelength give neutral atomic hydrogen gas content of about 17 per cent by mass in the irregular galaxies; the *Sc* galaxies have about 8 per cent, and *Sb* systems like the Andromeda nebula, have only 1 or 2 per cent. This same figure holds for our Milky Way, so that in the *Sb* systems

FIGURE 7. A young spiral nebula, M 101, of type *Sc*. The arms are clearly defined and contain stars, or groups of stars, of very high luminosity. The central nuclear region is very small. Population I dominates.



NGC 1201 Type S0

NGC 2841 Type Sb

NGC 2811 Type Sa

NGC 3031 M81 Type Sb

NGC 488 Type Sab

NGC 628 M74 Type Sc

FIGURE 8. Montage of photographs of spiral nebulae, starting at lower right with the youngest, of type *Sc*, with pronounced spiral arms, gas, and high-luminosity stars, and ending at upper left with type *SO*, in which the gas is largely exhausted and the arms nearly absent. Note the growth of the central nuclear regions, from nearly a point, to a large ellipsoidal mass of relatively faint, red Population II giants. The inclination of the objects to the line of sight is variable, and some systems are highly foreshortened.

the star-formation process (which is very density-dependent) should be nearly finished.

SUPER EXPLOSIONS

The sun has intense flares releasing x-rays and soft cosmic rays; some stars have larger flares. The magnetic stars have enormous general magnetic fields which are believed to accelerate large numbers of particles to high energies. The cosmic rays pervade the Galaxy, the most energetic being possibly too rigid even to be contained by the Galaxy. There is a galactic magnetic field, and other galaxies have a field, as shown by the Faraday rotation of the plane of polarization. Our Galaxy has some high-energy electrons, and the radio sources have enormous total energy content in the magnetic field and in the relativistic electrons. Until now we have looked for the source of these energetic events in the realm of magnetohydrodynamics. Whether such a normal explanation of the abnormal, high-energy physics end of astrophysical phenomena is even possible, not to say plausible, is not yet known. Various Fermi or betatron acceleration processes in space have been suggested.

Stars are subject to nuclear and possibly gravitational catastrophes. It is not impossible that such giant explosions are the source of the nearly universal presence of high-energy particles. But even an explosion at a billion degrees travels with a speed of only 5000 km/sec., far below relativistic velocities. Nuclear reactions yield only a few million electron volts per reaction, also far from relativistic. Only extreme gravitational collapse gives much promise. If the sun shrank to a kilometer radius, a proton would drop into a gravitational potential well one billion volts deep, so that implosion-explosion events become interesting. But these involve a density of 10^{17} gm/cm^3 so that nuclear forces of short range and general relativity become an important feature of the existence of such objects. At another extreme, if stars of moderate density and very large mass, 10^6 to 10^9 $M\odot$ could be formed, gravitational collapse could also supply enormous amounts of energy. Perhaps an appreciable fraction of it could be converted into kinetic energy, magnetic fields, and high-energy electrons in an explosion.

The most exciting developments of recent years have been in the study of extragalactic nebulae by means of radio telescopes. Many identified radio sources are intrinsically weak, like our own Galaxy, but others are enormously strong. The strong sources emit 10^{44} to 10^{45} ergs/sec. in radio wavelengths only; the quasi-stellar radio source 3C48 emits 2×10^{45} ergs/sec. by high-energy processes, from radio frequencies to near ultraviolet. Relativistic electrons of energies up to 100 Bev spiralling in magnetic fields produce the so-called synchrotron emission. Minimum lifetimes of the radio galaxies are several hundred thousand years, so that the energy emitted is up to 10^{58} ergs, equivalent to 10^7 supernovae, or 10^7 times the total nuclear energy of the sun over its entire lifetime. The total relativistic electron and magnetic field energy content required to explain this radiation is at least 10^{60} to 10^{61} ergs in the radio galaxies.

Some type of explosion is required. The nearby, relatively weak radio source, M82, is shown in Figure 9, where the red hydrogen line, H alpha, shows material expelled out of the poles of a peculiar galaxy. M87, otherwise a normal giant EO elliptical galaxy, has a jet of polarized, synchrotron light near its nucleus, connected with a large radio source (Figure 10). The jet reappears in 3C273, a relatively bright star with a hydrogen emission-line spectrum redshifted by 0.16. This is a quasi-stellar radio source; if the jet was expelled with nearly the velocity of light, it is 200,000 years old; the visual luminosity of 3C273 is 3×10^{46} ergs/sec., more luminous than any other galaxy by a factor of twenty. To show the enormous extra spatial penetration now obtainable with the 200-in. Hale reflector, Figure 11 shows the spectrum of 3C295, a very strong radio galaxy at the largest known redshift, 0.46, and in contrast, Figure 12 shows a quasi-stellar radio source 3C48 at nearly the same redshift, 0.37. (A redshift of 0.37 means $\lambda/\lambda_0 = 1.37$.) Only a single weak emission is found in 3C295, while 3C48 which is nearly 100 times brighter, has a rich spectrum, including a line of Mg II, from the normally unobservable ultraviolet, at $\lambda 2800$. There is little doubt that the quasi-stellar radio sources already known have enormous redshifts and will carry us out close to the boundaries of observability.

The nature of these extraordinarily luminous objects is not clear. Their redshifts are quite certain, since seven spectral lines in each quasi-stellar radio source give quite accordant shifts. It seems impossible to explain the redshifts by intense gravitational fields, because of the large volume of gas required to produce the observed forbidden emission lines. No galaxies, or clusters of galaxies, are associated with the apparently stellar images. The sources have small angular diameters, from radio and optical measures — less than a second of arc. If the

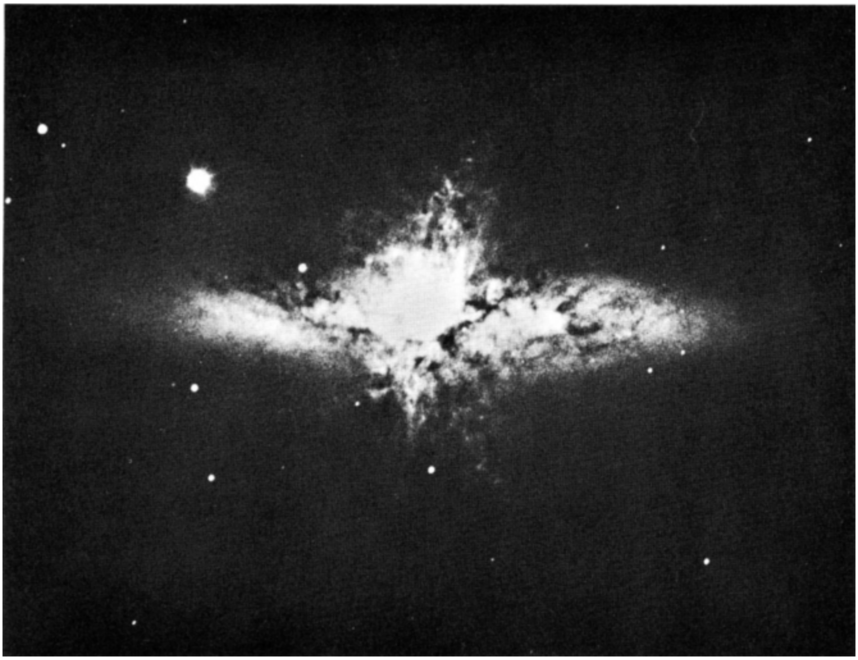


FIGURE 9. M82 is a peculiar galaxy with large amounts of dust. In this composite photograph, the normal picture of the stellar content has been nearly canceled out by superposition of a negative and positive. The photograph shows the H-alpha emission of exploding gas clouds running in a vertical direction along the minor axis.

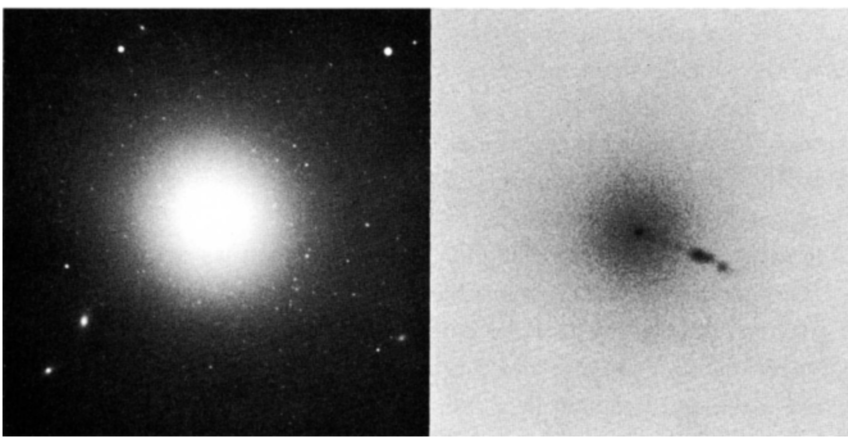


FIGURE 10. M87 is a giant elliptical galaxy with only bright globular clusters resolved (left). The inner region is shown with greater enlargement as a negative print, on the right. It contains a jet of highly polarized, reddish light, probably of synchrotron origin.

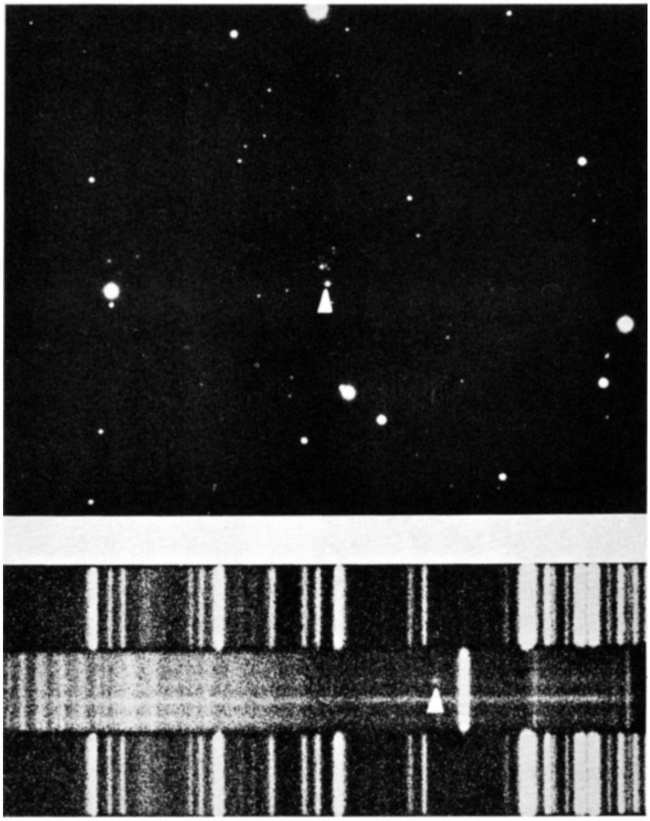


FIGURE 11. The radio source 3C295, the most distant known galaxy, is a member of a faint cluster of galaxies. Its redshift is such that $\lambda/\lambda_o = 1.46$. The spectrum (below) obtained by Minkowski shows only one weak emission, probably $\lambda 3727$ of 0 II; the lines running completely across the spectrum are terrestrial airglow in origin (positive).

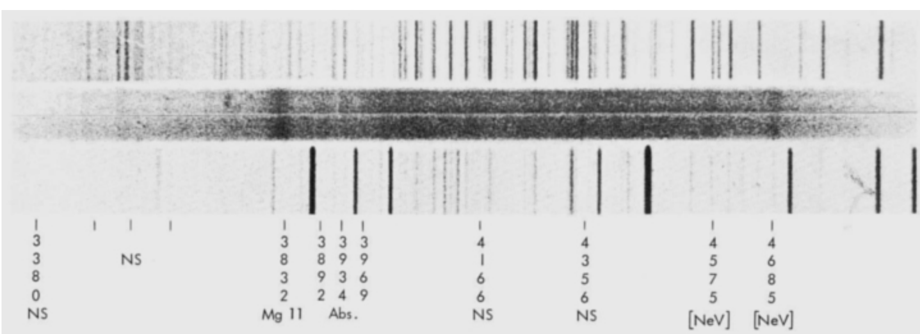


FIGURE 12. The radio source 3C48, the quasi-stellar radio source, has $\lambda/\lambda_o = 1.37$. Two negatives of the spectrum of this relatively bright object are mounted together to show the reality of many diffuse emission lines. Those labeled NS are terrestrial airglow. Objects like 3C48 will be observable to $\lambda/\lambda_o > 2$.

redshifts are cosmological in origin, the distances of the objects are 2 to 4 billion light years, so that this upper limit means only that the diameters are less than 10,000 light years. No traces of stellar spectra are seen. Analysis of the spectra provides an acceptable range of space density of gas from 10^4 to 10^8 electrons per cm^3, corresponding to masses of 10^9 to 10^4 $M \odot$ and radii of 300 to less than one light year. The emission lines are broad, indicating internal or expansion velocities of 1500 km/sec. If they are massive and large, the internal kinetic energy is sufficient to maintain their luminosity for 10^5 years, if an efficient, unknown mechanism converts kinetic energy into magnetic field and relativistic electrons for the synchrotron process. But if they are small, they contain energy for only a few hundred years. Their light is almost certainly variable, suggesting small size. If the small size is proved correct, we require an unknown, enormous supply of energy from a very small volume. Chain reactions of supernovae would barely suffice, even in a very densely crowded group of stars. Thermal energies are 10 electron volts per nucleon; to maintain the luminosity for 500 years requires 30 kilovolts per nucleon, and for 200,000 years over 10 million volts per nucleon. Thus either an unknown new type of super supernova, an enormously massive star, or some unknown storehouse of energy is required to explain these great explosions. In the quasi-stellar radio sources like 3C48 or 3C273 we are not even certain whether the objects are in galaxies or are newly formed in intergalactic space. We do not know whether they represent new matter created by singularities in the geometry of space, or are gathered from the intergalactic background, or whether they are common events in old galaxies as are the ordinary intense radio sources. In any case, they promise to carry cosmological studies further and to provide exciting work for astronomers, theoretical physicists, and cosmologists.

The History of the Solar System

FRED L. WHIPPLE

INTRODUCTION

WHAT HAS HAPPENED in the past appears to be almost as vague and uncertain as what will happen in the future: the story depends upon the teller. Contrast the histories of a war as written by the winner and the loser, or the stories of a courtship as related by the bride, by the groom, or by friends. Note that observers of automobile accidents almost invariably disagree as to the actual sequence of events. And so it is, even in science. We can reconstruct the history of the solar system with little more confidence than we can predict its future. Actually, we possess only a fragmentary knowledge of the system today and have inadequate theoretical tools to deal with many of the physical processes that have taken place.

Since there are many contradictory arguments regarding the history of the solar system, this will also be a biased account stressing a few of the areas where there is a strong consensus, and presenting only part of the counterevidence where great uncertainties or unusually strong differences of opinion exist.

We must make one general assumption to avoid wallowing in the quicksands of sheer speculation (or metaphysics?); viz., that the laws of nature have remained unchanged for some five aeons (an aeon is defined as 10^9 years). To my knowledge only one measurable quantity has certainly remained constant over some three aeons; i.e., the range of alpha particles from specific radioactive atoms in mica, as evidenced in pleochroic halos[1] (Figure 1). As we shall see, today's measured ages of the earth and meteorites agree well, although determined from different radioactive processes, thus indicating that certain relationships among the constants of radioactivity have not changed. There is no similar evidence to substantiate the assumption that such fundamental quantities as the

---

FRED L. WHIPPLE Astrophysical Observatory, Smithsonian Institution, Cambridge, Massachusetts

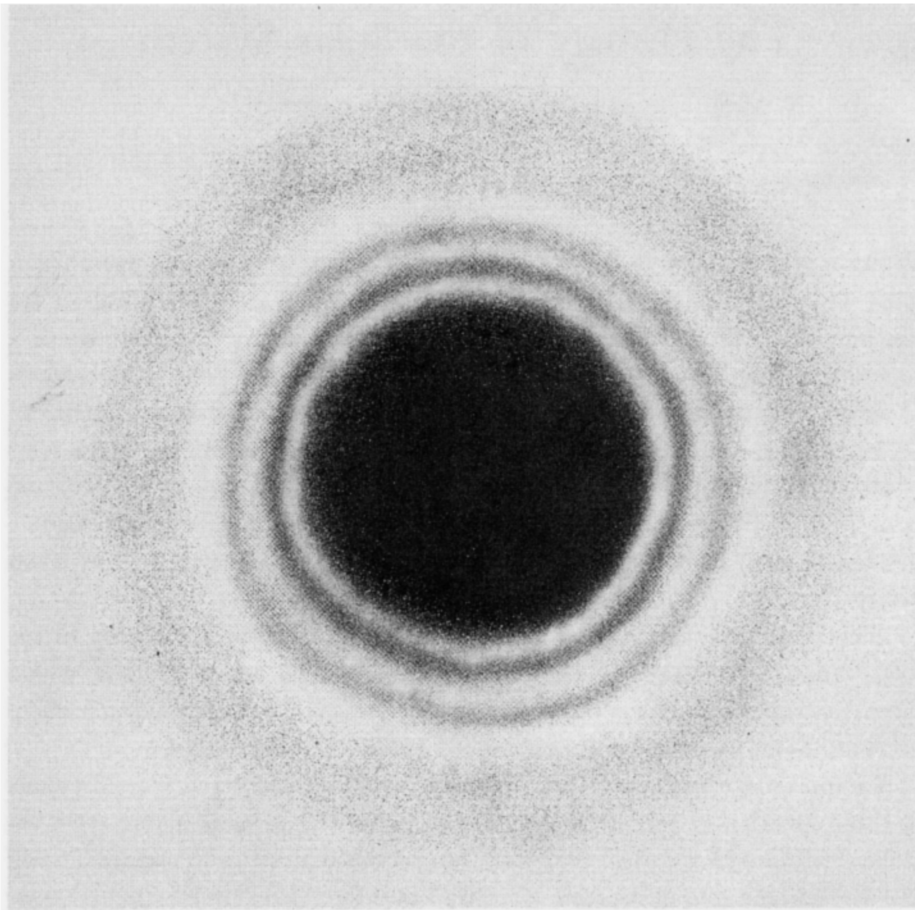


FIGURE 1. A pleochroic halo.

velocity of light, the constant of gravitation,* or other physical constants have remained unchanged during the past five aeons.

HISTORY OF THE EARTH

Assuming, however, that all is well with the physical laws, we can determine an age of the earth: that is to say, the time interval since the earth became something like its present self. At least we think we can. Figure 2 shows a few estimates of the age of the earth. It begins with the adopted biblical value, the smallest value, yet the one that has been believed for the longest time. Following are values determined by Helmholtz from the rate of the contraction of the sun, by Kelvin from heat conductivity, by Joly from the transport of salt to the oceans, by Holmes from radioactivity measures, and by recent investigators who employed a number of techniques involving radioactivity. On the average, the "age" of the earth has been doubling every 15 years for the past 3 centuries; the rate, perhaps, has been somewhat faster during the past century. There are some notable exceptions to these more classical estimates. Some Eastern philosophies postulate much greater ages. In the West, James Hutton, a geologist of the late 18th century, could see no evidence of a beginning or an end to geological processes. Perhaps these estimates best illustrate my first point that the past is highly variable.

Actually there is some real basis for thinking that the numbers are now converging. Other lecturers have discussed the facts of radioactivity and nucleogenesis, so I list in Table I some of the major processes of radioactivity that are useful in determining the age of the earth and of meteorites; included are certain end-products and the half-lives, or the intervals of time in which half of the atoms spontaneously decay. By measuring the parent atom and its daughter decay products *in a sample for which both are preserved* we can now determine the length of time these atoms have been contained in the sample. The analysis is not so straightforward should the material contain some of the decay products initially. Thus age determinations by the strontium-rubidium method or by the lead produced from uranium and thorium become too complicated for our exposition. The gases helium and argon, of course, can leak out of some kinds of rocks or

* Note that P. Pochado and M. Schwarzschild (*Astrophysical J.*, 1964, vol. 139: 587) find for the sun that G cannot have varied more rapidly than by $(\text{time})^{-0.2}$.

meteorites and oftentimes ages calculated for them are less than the time since the material became solid.

Table II lists ages determined for the earth and meteorites. Meteorites, of course, are stones and irons that have fallen from the sky. We have ample evidence that they belong to the solar system. No earth rocks yet found have remained cool and solid for more than three aeons. The lead-lead method, however, measures the age since the earth was assembled, regardless of thermal processes, mixing, and loss of gases. Claire Patterson[2] finds an age of 4.55 aeons by this method. Many of the meteorites solidified 4.0 to 4.4 aeons ago[3] but the age since their major consolidation, as measured by the lead-lead method, is in excellent agreement with that of the earth. Thus, a mean value of the age of the earth and meteorites appears to be within a few hundred million years of 4.6 aeons. William A. Fowler,[4] fortunately, synthesizes the elements in time for the earth to appear. There is, however, a real possibility not to be ignored that the lead-lead method may err appreciably; thus the consolidation of the earth and the parent bodies of the meteorites may have occurred earlier than we now believe. Remember Figure 2.

NEWTONIAN FACTS

Until about the beginning of this century, our knowledge of the solar system consisted almost entirely of what I shall call *Newtonian facts,* or those deductions that could be drawn from geometrical observations interpreted according to Newton's laws of motion and gravitation. A central star, the sun, containing more than 99 per cent of the total mass, controls gravitationally the motion of nine planets, all of which move in nearly circular orbits close to a common plane and in a common direction. Their distances from the sun follow

Table I. RADIOACTIVE ATOMS

Atom	Decay products	Half-life
U^{238}	$Pb^{206} + 8\ He^4$	4.51 aeons
U^{235}	$Pb^{207} + 7\ He^4$	0.71 aeons
Th^{232}	$Pb^{208} + 6\ He^4$	13.9 aeons
Rb^{87}	Sr^{87}	46 aeons
K^{40} (11 per cent)	Ar^{40}	1.25 aeons

Table II. AGES (IN AEONS, 10^9 YR.)

Method	Earth	Meteorites
U,Th – He	~3 max	4.0 (peak)
K – Ar	~3 max	4.3 (peak)
Rb – Sr	~3 max	4.4
Lead-lead	4.55	4.6 ± 0.2

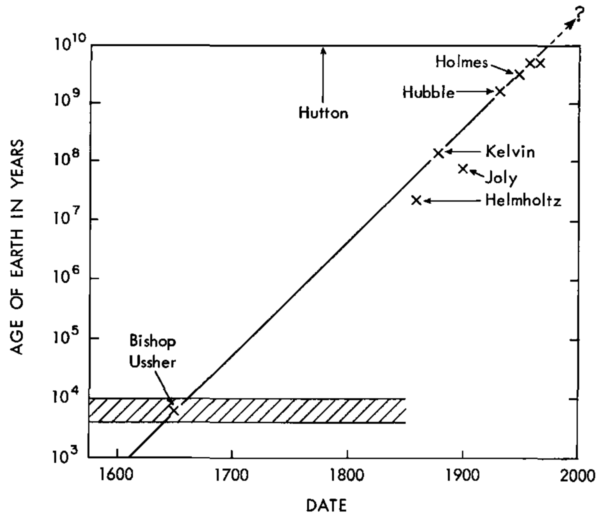


FIGURE 2. Dated estimates of the age of the earth.

a simple empirical relationship, the Titius-Bode law, as shown in Table III. This so-called law requires a hiatus between Mars and Jupiter to account for a huge number of minor planets and asteroids where the fourth planet should move. These asteroids, together comprising perhaps a thousandth of the earth's mass, move in the direct motion common to the planets, but with somewhat greater orbital eccentricities and inclinations to the common plane. They are almost certainly the source of the meteorites. Note that the "law" fails for Neptune.

Beyond the reaches of the planetary system, more than 50 astronomical units from the sun (i.e., 50 times the Earth's distance), we have reason to believe a cloud of comets, nearly in random motion, extends out to distances greater than 100,000 a.u., still gravitationally a part of the solar system. Jan Oort[5] estimates that there is a total of some 10^{11} comets containing possibly one earth's mass. A few are randomly disturbed by the passing stars and successively by Jupiter to become visible for a few revolutions through the inner reaches of the solar system. Most astronomers seem to agree with my theory[6] that the nuclei of comets are fundamentally frozen ices and dirt, that is, the expected cosmic material that can freeze at less than 100°K. We have learned from photographic meteor studies that ordinary meteors, or shooting stars, arise from cometary debris.[7] The incoming bodies are fragile in structure, not solid like most meteorites.[8]

The common motion of the planets is repeated in their rotations and in the motions of a considerable number of natural satellites, particularly around the great planets, Jupiter and Saturn. But there are exceptions. Uranus and its satellites are tilted a little more than 90 degrees to the common plane, while a few of the satellites have retrograde motions; viz., the outer four of Jupiter's twelve; the outermost of Saturn's nine; and the inner and larger of Neptune's pair. Radar measurements[9] suggest that Venus, almost the counterpart of the Earth, may also rotate (slowly) in a retrograde sense. Mercury keeps the same face toward the sun as does the moon towards the earth, presumably an effect of tidal friction.

The giant planets rotate very rapidly on their axes with periods of the order of 10 hours, while Mars rotates at about the earth's rate and Pluto, which possibly may not be a planet at all, in about 6 days. The sun rotates directly, but its equator is tilted 7 degrees to the plane of the planets as is the plane of Mercury's orbit.

The extremely slow rotation of the sun in 25 days presents an anomaly of enormous significance to our historic interpretation. Within a single system, Newtonian mechanics conserves an important quantity besides energy; viz., angular momentun. For a body in circular motion around the center of gravity, angular momentum is the product of velocity, mass, and distance. The sum of this quantity over all the masses in the system cannot be changed by any internal

Table III. The Titius-Bode Law

Planet	Basis	Sum/10	Distance *a.u.*
Mercury	4 + 0	0.4	0.39
Venus	4 + 3	0.7	0.72
Earth	4 + 6	1.0	1.00
Mars	4 + 12	1.6	1.52
Asteroids	4 + 24	2.8	–
Jupiter	4 + 48	5.2	5.20
Saturn	4 + 96	10.0	9.54
Uranus	4 + 192	19.6	19.18
Neptune	–	–	30.06
Pluto	4 + 384	38.8	39.52

Table IV.
Mass and Angular Momentum

Material	Mass *per cent*	Angular momentum *per cent*
Sun	99.865	0.5
Terrestrial planets	0.0006	0.2
Giant planets	0.134	99.3

forces, but only by external transfer to material outside the system. The sun, amazingly enough, carries less than half a per cent of the angular momentum of the solar system, although it contains more than 99 per cent of the mass (Table IV). This peculiar distribution of angular momentum in the solar system proved a stumbling block to all early evolution theories.

No *likely* method of accumulating planetary masses about the sun, or of condensing the sun and the planetary masses from a larger aggregate, can result in a slowly rotating central mass, if we base our theory on classical dynamics and hydrodynamics. Hence, the well known early theories[10] formulated by Immanuel Kant and Pierre Simon Laplace in the late 18th century, even when modernized, cannot account for the sun's slow rotation. Figure 3 shows an artist's concept of Laplace's Nebular Hypothesis, in which the rotating and collapsing sun leaves off successive rings of material at the distance of the present planets. Even granting with Laplace that such a condensing sun can, indeed, leave behind rings or discoids of material in rotation, we cannot escape by any process of hydrodynamics from the conclusion that the finally condensed sun, comprising most of the mass, will be rotating very rapidly with a period of 10 hours or less.

Yet all old stars like the sun are rotating slowly, even though new stars often show extremely rapid rates of rotation. Some force not included in classical dynamics and hydrodynamics is clearly at work. To explain the anomaly, a number of investigators following Buffon in 1745 have postulated that the sun was struck by an external body (Buffon called it a comet) tearing out and spinning the mass necessary to produce the planets. Such theories, including passing or colliding stars and even a destroyed solar companion, have all failed when subjected to the detailed application of dynamical theory. Sir James Jeans' tidal theory is illustrated in Figure 4. As a step beyond Newtonian facts, Lyman Spitzer[11] has shown that stellar (solar) gas at a temperature of a million or more degrees within the sun or companion star and under high gravitational pressure would, upon violent release, explode far too rapidly for condensation into planetary masses or for accumulation into a disk about the sun.

C. Von Weizsäcker[12] applied modern turbulence theory to transfer angular momentum from the sun to a Laplacian nebular disk (Figure 5) but the process stops when the velocity at the equator becomes comparable with the circular orbital velocity. The sun is left rotating with a period of only a few hours. Hence, turbulence offers no solution.

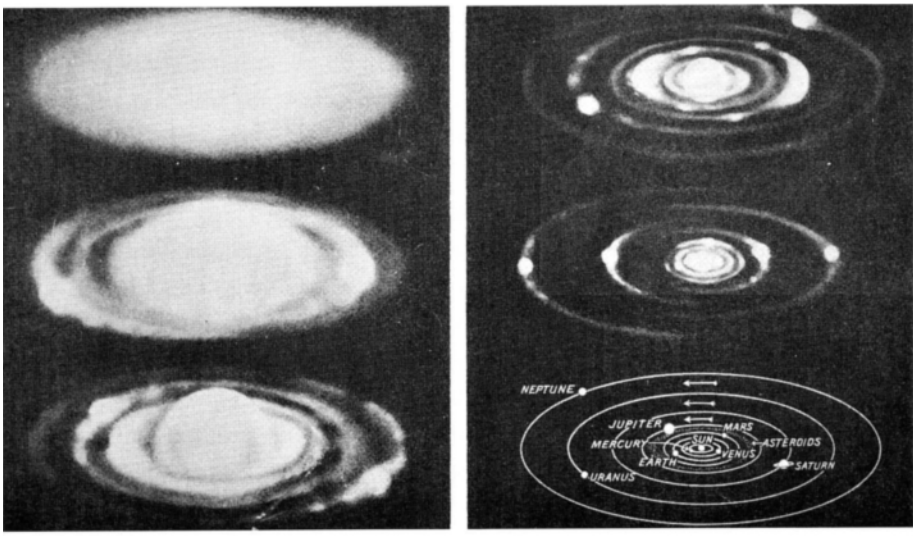


FIGURE 3. Drawings of Laplace's nebular hypothesis.

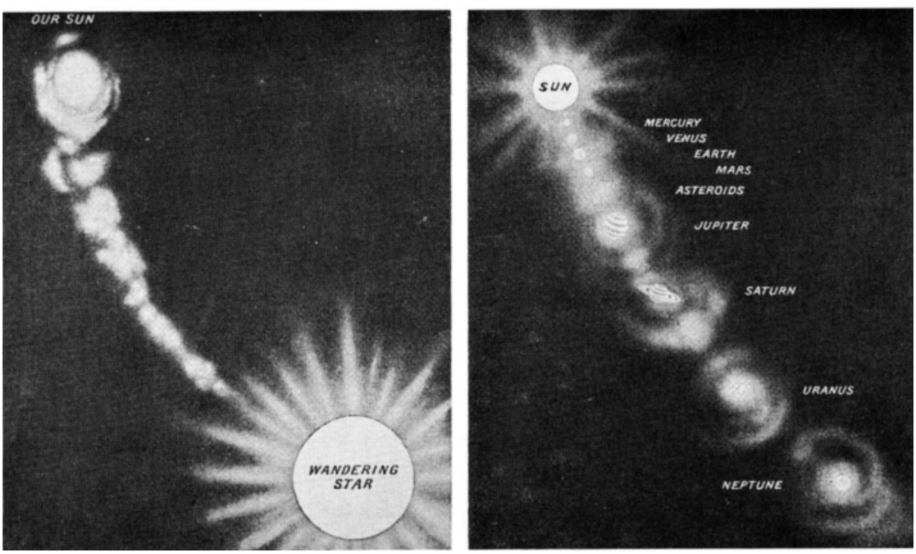


FIGURE 4. Sir James Jeans' tidal theory.

MAGNETOHYDRODYNAMICS

Today the problem of angular momentum appears to have a likely solution only when we postulate the action of *magnetohydrodynamic* processes. Their study is as difficult and important as the name is complicated. Magnetohydrodynamics deals with processes in hot, ionized gases where electrical currents and magnetic fields have energies comparable to those of normal gas motions, pressure, heat, and thermodynamic interactions. Where there is a magnetic field in space, charged particles can move easily *along* the field lines but can move *across* them only with difficulty. Thus the magnetic lines of the earth's field hold the extremely energetic charged particles of the Van Allen belts, as satellite astronomy has recently demonstrated. When such fields are produced within the gas itself by a flow of the charged gas particles, all the theoretical complexities of hydrodynamics are multiplied by the additional comparable complexities of electromagnetic effects.

In 1912 K. Birkeland[13] first suggested that a strong solar magnetic field might play a major role in planetary formation. H. P. Berlage[14] in subsequent decades investigated various electromagnetic processes, as did H. Alfvén[15] in 1942. Their theories, however, all required sets of physical processes too specialized to be acceptable. In 1948 D. ter Haar[16] suggested a method whereby a strong magnetic field in the early, rapidly rotating sun might transfer angular momentum to the interstellar medium of gas. In 1957 Donald H. Menzel,[17] and in 1960 Fred Hoyle,[18] suggested that ionized gas may have been thrown from the collapsing sun by magnetohydrodynamic forces to slow its rotation and remove gases from the system. In Menzel's theory the electromagnetic forces of an electric current produced a Saturn-like protrusion of the solar equator. These same forces restrained the gases of the expanding ring from evaporating into space. In Hoyle's theory the sun contracted rapidly to about the dimensions of Mercury's present orbit; at that time its rotation and increasing temperature produced strong magnetic fields near the surface. These fields produced "wiry" rigidity in the ionized gas, which caused the rotating sun to drag behind it the outermost gases and wind up the magnetic lines of force in discoids about the sun (Figures 6 and 7). The angular momentum of the sun was then transferred to this gaseous discoid, slowly carrying the gases outward from the sun in the plane of rotation.

Hoyle has calculated that about 1 per cent of the solar mass would be adequate for this process and that as the material cooled, the outgoing gas would carry with it the solid planetesimals which were below about 10 meters' radius. The solids and some of the gases would collect into the various planets. Today, whether or not we accept these detailed processes, we must of necessity call on some magnetohydrodynamic process for transferring angular momentum from the early rotating sun.

CHEMISTRY AND PHYSICS

The factual basis for our history has been richly increased this past century by the developments in chemistry and physics, coupled with the concomitant advances in engineering and observational techniques. Spectroscopy enables us to determine much about the composition and character of stellar and planetary atmospheres.[19] Thus, ammonia, methane, and hydrogen appear in the spectra of Jupiter and Saturn, while a considerable percentage of helium can be demonstrated by indirect means. Methane appears in the spectra of Uranus and Neptune but ammonia is apparently frozen out. Correspondingly, carbon dioxide occurs

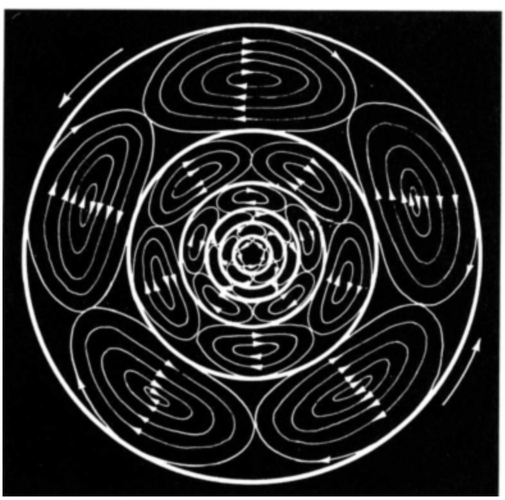


FIGURE 5. C. von Weizsäcker's concept of turbulent eddies in the solar nebula.

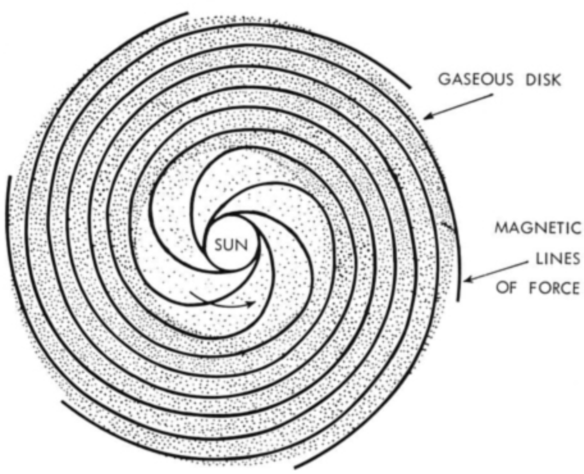


FIGURE 6. Magnetic lines in the solar nebula according to F. Hoyle

in the atmospheres of Mars and Venus and water vapor in Mars. Most of the known stable elements are observed in the sun's atmosphere and their relative abundances have been determined.

As you have already heard, progress in the theory of stellar interiors and energy generation provides excellent knowledge of the distribution of temperature and pressure within the sun. Correspondingly, high-pressure and solid-state physics have provided considerable information concerning the internal structure of the giant planets. R. Wildt,[20] W. H. Ramsey,[21] and W. C. DeMarcus[22] have shown that Jupiter must consist primarily of hydrogen and is almost the maximum size for a cold planet. Were it much more massive (Figure 8), its central material would become physically *degenerate* (i.e., very much denser) and it would be smaller; if less massive, it would also be smaller. Saturn, with a much lower mean density — 0.7 *versus* 1.3 that of water — actually contains a larger fraction of the heavier elements.

Seismic data and theoretical studies of the earth's interior (Figure 9) show that it, on the other hand, may contain an iron core, possibly liquid. This supposition is strongly supported by the occurrence of many iron-nickel meteorites (Figure 10), which suggests a sequence of events: (*a*) the growth of two or more sizable asteroids whose interiors were or became heated, (*b*) the gravitational separation of the denser melted iron towards the centers, leaving the lighter silicates in the mantles, (*c*) the breakup of the asteroids presumably by collisions after a huge interval of cooling, (*d*) the final encounters of a few

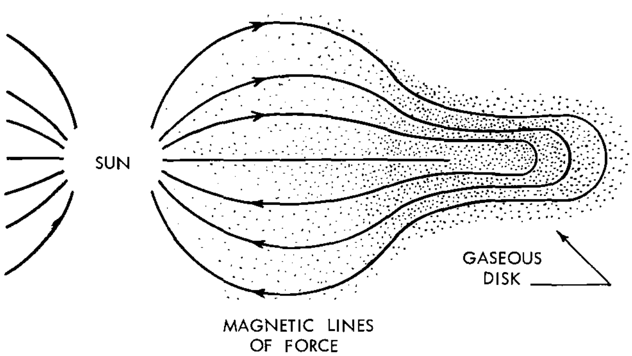


FIGURE 7. Edge view of Hoyle's magnetic lines.

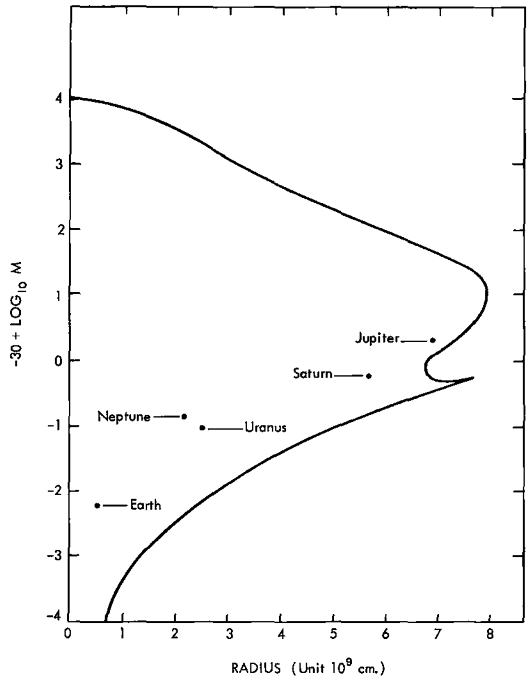


FIGURE 8. The radii of the planets as compared to hydrogen spheres of the same mass, after W. C. DeMarcus.

fragments with the earth. This evidence once strengthened the intuitive feeling that the earth must have been born as a hot gaseous body and cooled over the aeons. We shall see that the chemical evidence, adduced particularly by Harold C. Urey,[23] disproves this hypothesis.

Among the terrestrial planets we have an interesting decrease in mean density from the sun outward as shown in Table V. The correction to the uncompressed state follows Urey's calculations. Mercury must certainly represent a much higher concentration of iron than the earth and Mars.

Urey[24] has repeatedly pointed out that the moon presents a puzzle of major significance; it is less dense than the stony meteorites and comparable in density to the rocks of the earth's upper mantle. How could it have been formed from the same materials as the earth and at the same time and place? Furthermore, he stresses that its irregularities and surface features indicate that it has never been thoroughly melted. With regard to its density, note too that the satellites, where

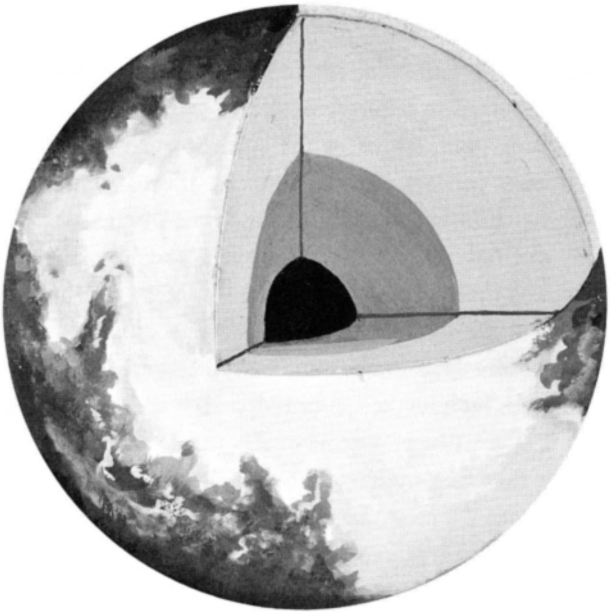


FIGURE 9. Interior of the earth showing the core; suspected to be nickel-iron.

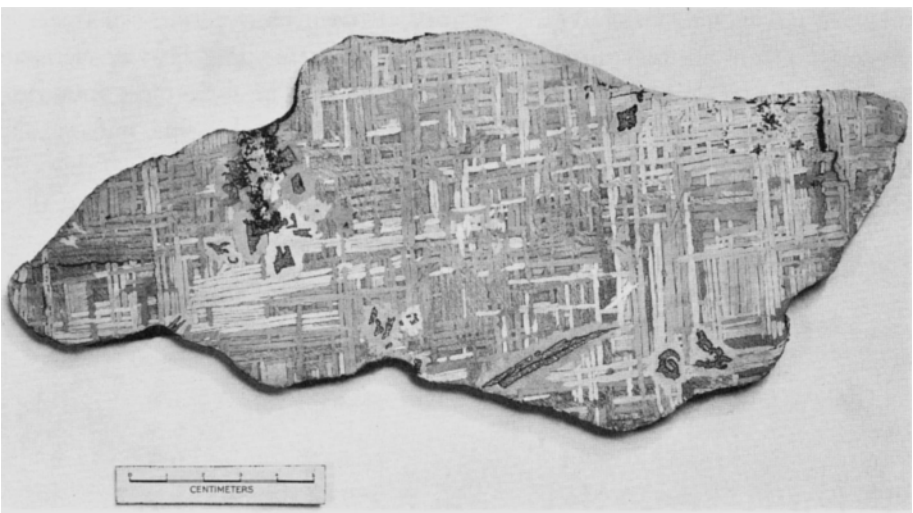


FIGURE 10. The Goose Lake iron meteorite, a polished and etched section showing the Widmänstatten figures.

data are available, become less dense as we go outward through the solar system, and also as we go outward in Jupiter's system (Table VI).

Measures of atomic abundances in the sun show that hydrogen is an overwhelming constituent, perhaps 60 per cent or more, helium perhaps half as abundant by weight, lithium, beryllium, and boron practically absent, while carbon, nitrogen, and oxygen constitute about 1 to 2 per cent. All the heavier elements comprise less than 1 per cent of the total mass. Beginning with carbon, nitrogen, and oxygen, the heavier elements in meteorites and in the earth show much the same abundances as in the sun, with a few notable exceptions, particularly the noble gases, which are essentially absent from the earth and most meteorites. Among the comets we have no detailed information concerning composition except that fragmentary compounds of carbon, nitrogen, oxygen, and hydrogen are detected in the spectra while the meteoric spectra show largely earthy elements such as iron, silicon, magnesium, sodium, calcium, etc.

Harrison Brown[25] pointed out that the elements could be divided into three natural groups in terms of abundance and physical characteristics. The most abundant elements, hydrogen and helium, remain gaseous at extremely low temperatures and constitute what I shall call the *gaseous group*. The compounds of carbon, nitrogen, and oxygen with hydrogen, such as ammonia (NH_3), methane (CH_4), and water (H_2O) vaporize at room temperature but freeze at moderately low temperatures, the *icy group*. The remaining heavier elements excluding the noble gases constitute the *earthy group*. The three most abundant, silicon, magnesium, and iron, either as compounds or as elements, remain solid to rather high temperatures, generally melting around 2000°K. Table VII illustrates the relative abundances of the three groups of elements and their

Table V. DENSITIES OF THE TERRESTRIAL PLANETS

	Actual (water = 1)	Uncompressed (water = 1)
Mercury	5.5	5.0
Venus	5.1	4.4(?)
Earth	5.52	4.4
Mars	4.0	3.7
Moon	3.34	3.31

Table VI. THE LARGE SATELLITES

Satellite	Diameter	Density (water = 1)
Moon (E)	1.00	3.3
Io (J)	0.99	3.1
Europa (J)	0.86	3.0
Ganymede (J)	1.28	1.7
Callisto (J)	1.10	1.4
Titan (S)	1.38	2.3
Triton (N)	1.06	?

melting points. The data are based upon determinations by Goldberg, Müller, and Aller[26] for the sun, and a compilation by H. E. Suess.[27]

We can estimate the relative abundances of the earthy, icy, and gaseous materials in the planets and in the comets, as shown in Table VIII. The earthy elements constitute essentially all of the mass of the terrestrial planets but rather small fractions of the giant planets and comets. I assume that comets consist entirely of the icy plus the earthy groups. Actually it appears that Jupiter and Saturn contain a few earth masses of earthy materials and Uranus and Neptune perhaps one or two each. The icy materials become increasingly significant as we go outward from the sun. Uranus and Neptune could be made almost entirely of comets. Jupiter and Saturn probably contain somewhat more icy material in absolute amount than Uranus and Neptune, even though the relative abundances are lower.

We may reasonably assume that the original source of building material was essentially the same for the sun as for the planets, whether the planets were derived from solar material or whether they were all condensed from the same gaseous clouds. The abundance of the elements strengthens this natural assumption even though there are some striking discrepancies. On this basis, the terrestrial planets, which contain only the earthy fraction, must have been derived from about 500 times their present mass. We see from Table IX that the original source material for the terrestrial planets was somewhat comparable to that for the giant planets. For the comets I assume 10 earth masses of original earthy and icy material. This allows for a 90 per cent loss during orbital changes by perturbations of the major planets to produce the huge cometary cloud, extending to 100,000 a.u. from the sun. Thus we see that if the condensation

Table VII. CLASSES OF MATERIAL

Material	Earthy	Icy	Gaseous
Elements	Si, Mg, Fe etc. plus O	C, N, O plus H (Ne?)	H, He
Mass available	1	4–7	300–600
Melting point	~2000°K	≤273°K	≤14°K

Table VIII. COMPOSITION OF THE PLANETS

Material	Earthy	Icy	Gaseous
Terrestrial planets	1.00	<0.01	0
Jupiter	<0.01	0.1	0.9
Saturn	0.01	0.3	0.7
Uranus	0.1	0.8	0.1
Neptune	0.2	0.7	0.1
Comets	0.15	0.85	0

process were highly efficient we require about 3 per cent of the present solar mass to provide the original material from which the planets and comets were derived. The efficiency of the collection processes surely must have been low so that the original source of material must have been at least one-tenth of the solar mass, an order of magnitude more than Hoyle assumes.

On this basis then, the original material from which the earth was formed constituted at least Jupiter's present mass of primordial material which, if it had been collected together, would surely have condensed into a planet much like Jupiter and remained as a giant planet. The same is even more certain for Uranus and Neptune. The difficulty of removing so much gas from large "protoplanets,"

Table IX. ORIGINAL PLANETARY MASS

Objects	Present mass (earth = 1)	Factor	Original material (sun = 1)
Terrestrial	1.9	500	0.0028
Jupiter	317	10	0.0095
Saturn	95	30	0.0086
Uranus and Neptune	32	75	0.0072
Comets	1	900	0.0027
Minimum original mass			0.0308

as postulated by Gerard P. Kuiper,[28] has invalidated his detailed theory of planetary formation. His contributions to the subject, however, are highly significant. We must find some process whereby the gaseous and icy materials were separated from the terrestrial planets or by which the earthy materials condensed, leaving behind the volatile materials at temperatures much like those that obtain on the earth today, or higher. At the present time a small body made of the icy materials will vaporize in the vacuum of space out beyond the distance of the asteroids.

Henry Norris Russell and Donald H. Menzel[29] many years ago showed that the earth's atmosphere had lost almost all its noble gases. Urey, by an extensive chemical study, has shown that the earth and the meteorites show significant losses of relatively *volatile* materials, compounds and elements more volatile than the element mercury. The icy elements were probably lost in the form of water, ammonia, and methane but some of the water, nitrogen, and carbon were preserved in compounds with the earthy elements. From this truly beautiful line of chemical evidence, which is too involved for our discussion, he visualizes that objects of some hundreds of kilometers in dimension accumulated at temperatures below O°C early in the history of the solar system. These primary objects were then heated variously, probably by internal radioactivity, to temperatures sometimes reaching 2000°C. They were then broken up by collisions and reaccumulated into the present asteroids, moon, and terrestrial planets. "During the process of breakup and reaccumulation, fractionization of the silicate and metal phases occurred, with a preferential loss of silicate phases in various degrees. This variable loss accounts for the variation in densities of the planets." Thus we now have a sound chemical basis for a planetesimal concept of planetary accretion, much more advanced and detailed but supporting some of the ideas in the well known theory presented by Chamberlain and Moulton early in this century.

For the earth, it is difficult to see how even the secondary planetesimals could have aggregated into such a large body with the volatiles so *completely* removed, particularly the heavy noble gases such as xenon and krypton. (Note that some meteorites contain appreciable quantities of these gases.) It seems very likely that the earth was in some fashion stripped of its atmosphere after formation and that the present atmosphere and oceans are all of secondary origin, probably derived from chemical compounds released by heating from within the earth, as the work of William Rubey[30] supports. The thin atmosphere of Mars and the absence of an appreciable one on Mercury present no problems. Because of their low surface gravities, atmospheres are easily lost. We still await more information about the atmosphere of Venus.

For the meteorites and the asteroids the historical problems are extremely difficult, as suggested by the complicated sequence of events required by Urey. It was long believed that two planets or more, possibly comparable to the earth in size, were formed in the asteroid belt to develop iron cores and finally, after collisions, form the present asteroids and the iron and stony meteorites. Conceivably evidence for two major parent asteroids can be derived from the discovery by Urey and Craig[31] that the chondritic (stony) meteorites divide into two classes with differing iron content and correlated abundance differences among the less abundant elements. This remarkable separation of the meteorites into two classes still remains unexplained by any chemical or physical theory.

In recent decades the presumed dimensions of the parent asteroidal bodies have continuously decreased from planetary to lunar size and recently, through the arguments of Goles, Fish, and Anders,[32] to relatively small asteroids of only a very few hundred kilometers in diameter. This theoretical shrinking arises from the difficulty of cooling large bodies within the time scale allotted, the order of a few hundred million years. Note from Table II that some of the meteorites have remained fairly cool and undifferentiated for more than 4.0 of our total of 4.6 aeons. The problem of maximum asteroid size is not yet settled. Almost no one nowadays maintains that the original asteroids were more than lunar size but various investigators advocate original dimensions from 300 to 3000 kilometers in diameter.

There is a wide difference of opinion as to the pressures required to produce the famous Widmannstätten figures (see Figure 10) in iron meteorites, somewhere in the range between values in the moon and in asteroids. Slow cooling is definitely required below a few hundred degrees centigrade but almost all theories allow for this. The diamonds occasionally found in certain meteorites point to a high-pressure, steady state in the original asteroid although a minority of investigators believe that they can be produced by intense shock. When a complete theoretical understanding of the production of diamonds and the Widmannstätten figures is reached, the problem of the maximum size of asteroids will be solved.

The moon enters the discussion because of its extremely low density and its irregularity in shape (more than 1 kilometer), which suggest that the moon was never completely molten. If the moon has not been thoroughly melted by radioactivity, how could 300-km diameter asteroids have been so melted? Their larger area per unit mass would surely accelerate radiative cooling and reduce the efficiency of heating processes.

We are not certain enough of the total radioactive content of the earth and meteorites to state definitely whether the known sources of radioactivity, particularly potassium,^{40} uranium, and thorium, could indeed melt the moon in 4.6 aeons. For smaller bodies this source of heat supply is too small unless they were originally hot, very near the melting point. An alternative would be to push back the date of formation so that the amount of K^{40} (with a half-life of 1.25 aeons) would have been more abundant, thus providing more heat. Remember that we are allowed only 4.6 minus 4.0 aeons, or 600 million years (or less) for the asteroids to heat and cool. Perhaps we should extrapolate our time-scale diagram (Figure 2) for another few years, gaining perhaps an aeon.

Analysis of meteorites[33] is today adding a wealth of information regarding the asteroids and the planetesimals that formed them. Of vital importance are the *chondrules,* small silicate droplets of dimension the order of 1 mm, present in and sometimes highly abundant in almost all stony meteorites (Figure 11). Many are glassy, indicating extremely quick cooling in a matter of minutes. In them the relative abundances of non-volatile substances fall closer to the observed solar abundances than in any other type of substance. Furthermore, they tend to contain metallic iron rather than oxidized iron, suggesting that they were formed in a reducing atmosphere, presumably hydrogen, as one would postulate in the original solar cloud. Their ages are great, indicating their formation during the early stages of planetary formation.

Hans E. Suess[34] and John A. Wood, Jr.,[35] have suggested that chondrules condensed directly in the primitive nebula. It is rather difficult, however, seriously to postulate conditions of sufficiently high pressure to make such direct condensation possible. As an alternative to this, Wood suggests that after having aggregated as small collections of dust, they may have been quickly melted and cooled by the passage of intense shock waves through the primitive nebula. Whatever the origin of the chondrules, no theory of the formation of the terrestrial planets can be complete without describing in detail the formation of so numerous a class of primitive objects.

NUCLEAR PHYSICS AND RADIOCHEMISTRY

In 1947 Harrison Brown[25] suggested that the time interval from nucleosynthesis of the elements to the formation of the asteroids might be determined if the daughter product of an extinct nuclear isotope could be detected in meteorites. A possibility for early planet heating is iodine,^{129} having a half-life of 17 million years. The daughter product is xenon,^{129} retained in a large planet or in a cool stone or iron. After considerable search, Reynolds[36] found this isotope with an excessive abundance in the Richardton chondrite and subsequent studies have shown that its abundance in other meteorites is highly variable with respect to

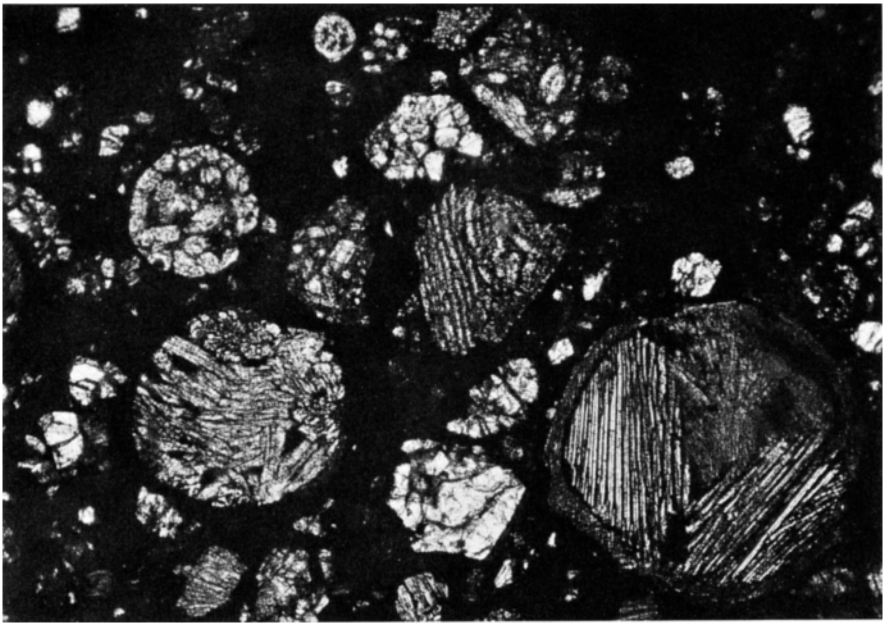


FIGURE 11. Section of the Tieschitz stony meteorite showing chondrules (diameters approximately 1 mm).

FIGURE 12. The Rosette Nebula in Monoceros showing dark clouds.

the other isotopes of xenon, sometimes being more than 3 times as abundant relatively as on the earth. This result suggested that a period of only 20 to 40 million years elapsed from the formation of the elements in stars or a supernova to the formation of the asteroids! Such a limitation on the time scale is exceedingly worrisome because it leaves no adequate time for cooling of the asteroids even if the contraction and planetary-forming processes could occur so surprisingly fast.

Another exciting puzzle in radiochemistry concerns the light elements deuterium, lithium, beryllium, and boron which, although not plentiful, are definitely present in meteorites and in the earth. On the sun they are nearly absent and should be, because they are destroyed at relatively low temperatures by thermonuclear reactions in stellar interiors. However, cosmic rays and less energetic particles can produce these isotopes in heavier material by the process of spallation, that is, by breaking atomic nuclei in violent collision. Thus Fowler, Greenstein, and Hoyle[37] have developed a theory to account for the presence of these elements, not as a consequence of thermonuclear reactions in the stars that provided the solar system gas, but by interaction with high-energy particles such as are now observed in solar flares. More frequent and more violent flares occur on young stars and may well have occurred on the young sun. Spallation alone, however, would have led to comparable ratios for the isotopes lithium^{6}/lithium,^{7} boron^{10}/boron,^{11} whereas these ratios are actually quite small for the earth and meteorites. The theory depends further on the ability of *slow* neutrons selectively to transmute lithium^{6} into tritium and boron^{10} into lithium.^{7} The spallation process, however, produces *fast* neutrons which will not reduce the Li^{6} and B^{10} to produce more $Li.^{7}$ In solid bodies containing hydrogen the neutrons can be slowed down to produce the desired result. Ice spheres of a few meters in dimension are precisely such bodies. Only a moderate fraction of the earthy material need be irradiated by solar flares and irradiation of too large a fraction will destroy certain isotopes such as gadolinium.^{57} Henri Mitler[38] has further shown that the planetesimals may vary over a wide range in size and need contain only a small percentage of water, eliminating the requirement that they be ice spheres.

The theory leads on to other remarkable conclusions: that the irradiation process can also produce, firstly, iodine^{129} that later decays into xenon,^{129} thus accounting for one puzzling anomaly; and, secondly, radioactive aluminum,^{26} of half-life 740,000 years, to heat the young asteroids.

FORMATION OF THE GIANT PLANETS AND THEIR SATELLITES

With these facts in mind I shall first attempt to reconstruct a possible historical sequence for the aggregation of the giant planets and their satellites. My picture leans heavily on the conceptual developments by A. G. W. Cameron,[39] who in turn leaned heavily on theories by other investigators including G. P. Kuiper,[28] F. Hoyle,[18] L. Mestel and L. Spitzer,[40] and E. Schatzman.[41]

Our starting point can begin no later than with a protosun of 1.2 or more solar masses extending over a volume with diameter perhaps 100,000 a.u. It is gravitationally unstable and about to collapse. It will be a cool ($\backsim 50°K$) and relatively dense interstellar cloud containing perhaps 1000 hydrogen atoms per cc and considerable dust. It may be associated with or a part of a much larger cloud (see Figure 12). There is some turbulence in the cloud and a mean rotation much like the rate of rotation of the galaxy with a period the order of 10^8 years. It has become gravitationally unstable possibly because of increased pressure applied externally; e.g., by the pressure of nearby supernova explosions, by large turbulent eddies in the Milky Way or in the associated larger cloud, or possibly by instabilities in the larger cloud.

Our protosolar system carries a weak magnetic field of perhaps 5×10^{-6} gauss (10^{-5} of the earth's field). There is probably some central condensation prior to the major collapse, the central temperature and density being appreciably higher than the mean. As the cloud collapses, the core probably picks up more rotation and the magnetic lines of force are twisted through the cloud. Dust in the cloud helps radiate away some of the energy of contraction.

Collapse in such a cloud requires some specific disposal mechanism for removing the potential energy of gravitation. Radiation to space may be slow or difficult for various reasons, including the possibility that hotter regions may exist in space nearby. Heating of the gas by compression provides a sink for some heat, but not much. We turn to the dissociation of hydrogen molecules as a major sink, which can be followed by the ionization of hydrogen and double ionization of helium, plus the smaller energy sinks involved in heating, dissociating, and ionizing the heavier atoms. Thus the collapse takes place at a far, far greater pace than would be required by Helmholtz' contraction concept, which depended mostly on radiation to remove gravitational energy. The time of collapse can be surprisingly short, perhaps only a few million years, not much longer than free fall to the center.

During collapse, Cameron finds that the cloud can transfer very little angular momentum to the nearby gas field, although the magnetic fields can be maintained. Cosmic rays and radioactive substances continuously produce the ions and electrons needed. In the early stages, the magnetic field lines soon become badly twisted. The total angular momentum of the cloud finally becomes a serious factor in collapse when the equatorial diameter has contracted to an uncertain value in the broad range between 1000 and 50 a.u. The energy sink provided by the ionization of hydrogen and helium is so great that continued rapid collapse would occur except for rotation. The protostar must shed rapidly rotating material at its equator, leaving a huge disk tens to hundreds of a.u. across. This is precisely the nebular disk as visualized by Laplace. The central temperatures of the protostar will be fairly high but the disk temperatures may be well below 2000°K at the moment the material is shed from the equator. The temperature will drop extremely fast because of radiation. The fall to room temperature at the present distance of Uranus may take place in the order of a year and to 10°K in perhaps a million years.

Thus contraction will certainly vaporize all of the icy material but may not vaporize all of the earthy material out to the present distance of Jupiter or Saturn. Here the total disk material, if compacted to the plane, would correspond in mass to about 10 meters of water. The disk, however, consisting mostly of hydrogen, would initially extend across the equatorial plane more than an a.u. at a temperature of 1000° and contract towards the disk's plane proportionally to the temperature. Perhaps one-tenth of a solar mass is spread out over this very large discoid beyond Jupiter's present orbit. Within less than a century the radiation cools this material to about 1000°K.

If not all the dust grains of earthy materials initially present in the gas cloud have been vaporized, the earthy and then the icy materials will quickly condense on any earthy nuclei left to produce what I shall call "cometesimals." This occurs at all distances beyond the neighborhood of Saturn until the nebular cloud becomes too thin.

Now we have the task of making the giant planets and getting rid of the large excess of hydrogen and helium in our nebular disk. Most of the gas, dust, and cometesimals rotate regularly in fairly circular orbits about the protosun, which is not yet hot enough to radiate appreciable energy. Some irregular gas globs or sheets are still falling in to disturb the uniformity.

We expect to eliminate the excess gas by tying the magnetic lines of force from the sun to the gas via the partial ionization produced by radioactive material, cosmic rays, and giant flares from the sun. The latter may play a major role in the momentum transfer. The sun at this time can be considered to have a real surface perhaps in the neighborhood of Mercury's present orbit, perhaps much farther out.

Jupiter and Saturn, because of their high gas content, must have accumulated much of their mass from the original nebular cloud. Urey[42] suggests that the temperature dropped so low ($\sim$4°K) in the region of Jupiter and Saturn *that hydrogen froze!* Thus the cometesimals there and at greater solar distances should have contained solid hydrogen. We do not know enough about the chemistry of comets to confirm this assumption. On the other hand, a temperature of 4°K is extremely low for an active violent region of space near a contracting star. Not only would one expect material to be falling in sporadically, but the general volume at great distances may contain early-type new stars and other contracting clouds. At an assumed low temperature of only 10°K, the vapor pressure of solid hydrogen has gone up by more than a million times and the possibility of freezing hydrogen has vanished (by a factor of 10^3).

It seems more likely that Jupiter and Saturn must have accumulated primarily from the gaseous cloud, perhaps by the process discussed by Kuiper. Change concentrations brought the mean density above the critical limit calculated by Roche (in 1850). The Roche limit is the minimum density at which a mass of gas could be gravitationally stable against the attraction of a central mass, here the protosun. At lower densities the tidal circulation would prevent the gas from collecting gravitationally, while at higher densities a stable gravitational unit would accumulate. It is probable that the continued existence of magnetic field lines spiraling in the disk tends to retard the accumulation of large gaseous masses. Our postulated nebular cloud is near the critical surface density. No theory can predict such details *a priori* but it can provide circumstances which might reasonably lead to such results.

At greater distances from the protosun than the region of Saturn the cometesimals become the building-blocks and rapidly collect into the planets Uranus and Neptune, besides contributing a comparable but relatively small fractional mass each to Jupiter and Saturn.

The direct rotation of Jupiter and Saturn should reflect the rotation of the cloud from which they condensed. A planet formed by the aggregation of small solid bodies should also rotate in the forward sense. It will overtake bodies in smaller orbits at aphelion, where they move slower than the planet and will also tend to collect them on the sunward side to produce forward rotation. Encounters with bodies in larger orbits will occur near perihelion and more frequently on the night side, to produce a similar result. Uranus must have swallowed a rather large cometesimal or protoplanet that tilted its plane of rotation.

The giant planets move in planes that never deviate much more than 1° from their common plane. This proves that they were formed by a process that involved average motions of the accreting matter rather precisely in the plane. The high inclination of Pluto's orbit (17°) may be accounted for by perturbations, since Pluto's perihelion falls within Neptune's orbit. Is Pluto really a large comet? Or is it a lost satellite of Neptune? Perhaps we will never know until we land a space probe on it.

I suspect that the Titius-Bode law, which does not hold well for the outer planets, states only that protoplanets collect material over larger areas at greater distances from the central gravitational mass. Possibly mutual perturbation of the planets tends eventually to smooth out the gross irregularities in the original orbital distribution.

Because of their satellite systems it is clear that during formation the giant planets all possessed great thin rings not dissimilar to the Laplacian disk about the sun (Figure 13). Probably magnetohydrodynamics played little part in their development. The giant planets are still rotating quite rapidly.

We find that the innermost satellites of Jupiter are nearly of the density corresponding to earthy material. This suggests two possibilities: (1) that Jupiter in its formation was first gaseous but collected about it a large discoid, something like Saturn's rings, first of earthy material and later, as the discoid grew, of captured cometesimals; or (2) that the temperatures in the massive disk close to Jupiter were high enough to have vaporized the ices for a relatively long period of time so that the inner moonlets were composed of earthy material. At greater distances from the planet the cooling permitted the ices to collect into sizable

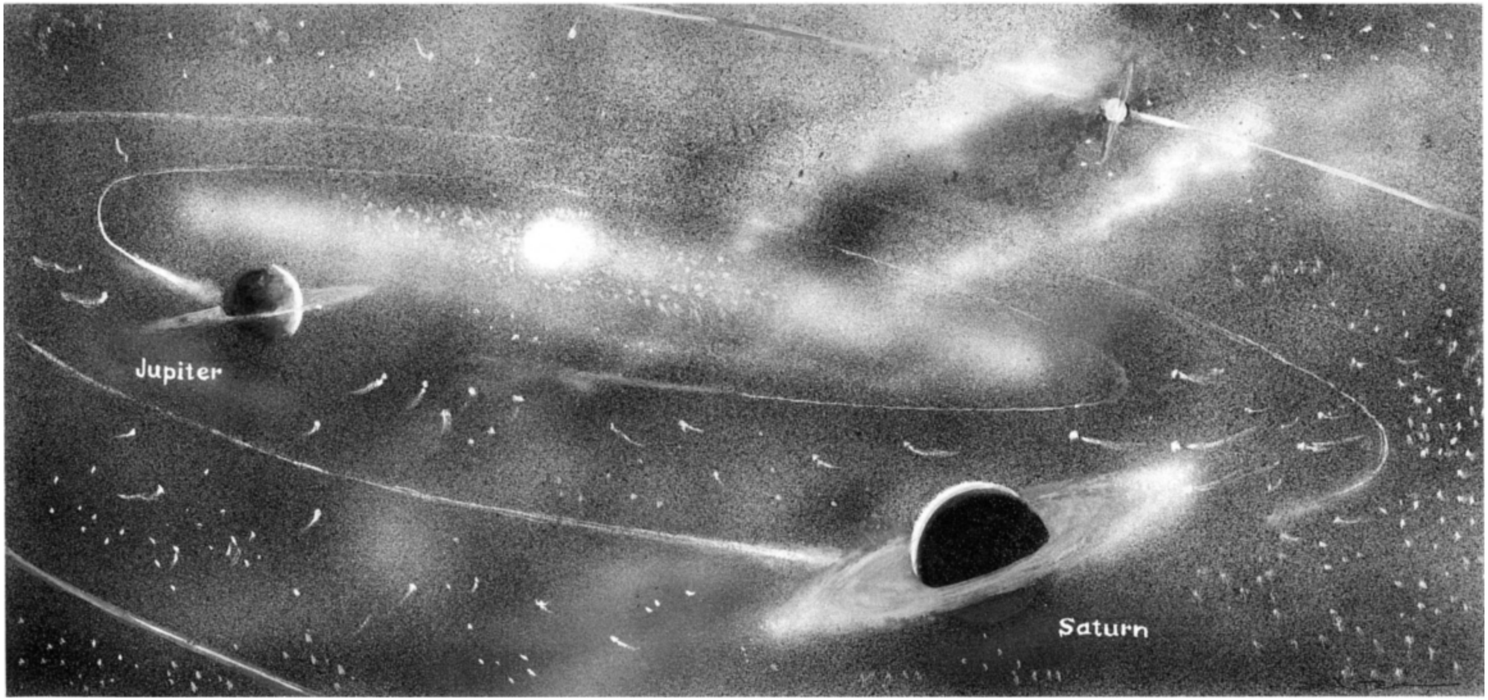


FIGURE 13. Drawing of planets during later stages of formation when comets were abundant.

satellites. Probably the original discoids about Jupiter and Saturn were even larger than the extent of their present satellite systems.

For Saturn we find that the satellite Titan is intermediate in density between icy and earthy materials, suggesting that Saturn developed an inner and earlier earthy ring that was subsequently increased by lower-density icy material to form Titan. Unfortunately we have no information concerning the densities of other satellites at the present time. The retrograde motion of the inner and massive Triton about Neptune, and the presence of only one other small satellite in a much larger direct orbit, point strongly to one conclusion: Triton was captured, and in the process destroyed most of the satellite discoid or the actual satellites. Lyttleton[43] suggests that Pluto was then ejected from the system.

THE ORIGIN OF THE COMETS

This reconstruction of the origin of the giant planets provides us with a multitude of comets found near the common plane of motion (Figure 14). Those that formed well beyond Neptune are presumably still there, as Cameron observes, forming a thin disk or belt in the deep freeze of outer space. They cannot contribute appreciably to the supply of comets that we now observe. They will be the subject of further discussion.

The great comet cloud extending out to perhaps 100,000 a.u. from the sun provides our present comets via the perturbations of passing stars. The comets of this cloud may well have originated near the plane of the disk in the region between Saturn and Neptune, having been perturbed into their present orbits by the gravitational actions of the planets. There are no stable orbits for small bodies in this region of space. Most of the cometesimals were probably captured by the planets to constitute the major fraction of Uranus and Neptune and a comparable mass each in Jupiter and Saturn. Some were perturbed into the sun and others thrown out to interstellar space. Still others decayed in the inner regions of the system as they do today, while rapid collisions eliminated many. The remainder were thrown into extremely elongated orbits to provide our comet supply.

A second alternative concerns the possible direct condensation of comets in small fragments of the original cloud before final collapse. McCrea[44] maintains that such an aggregation process is possible. If so, the comets we see today may contain representatives from both sources. The two sources could provide comets

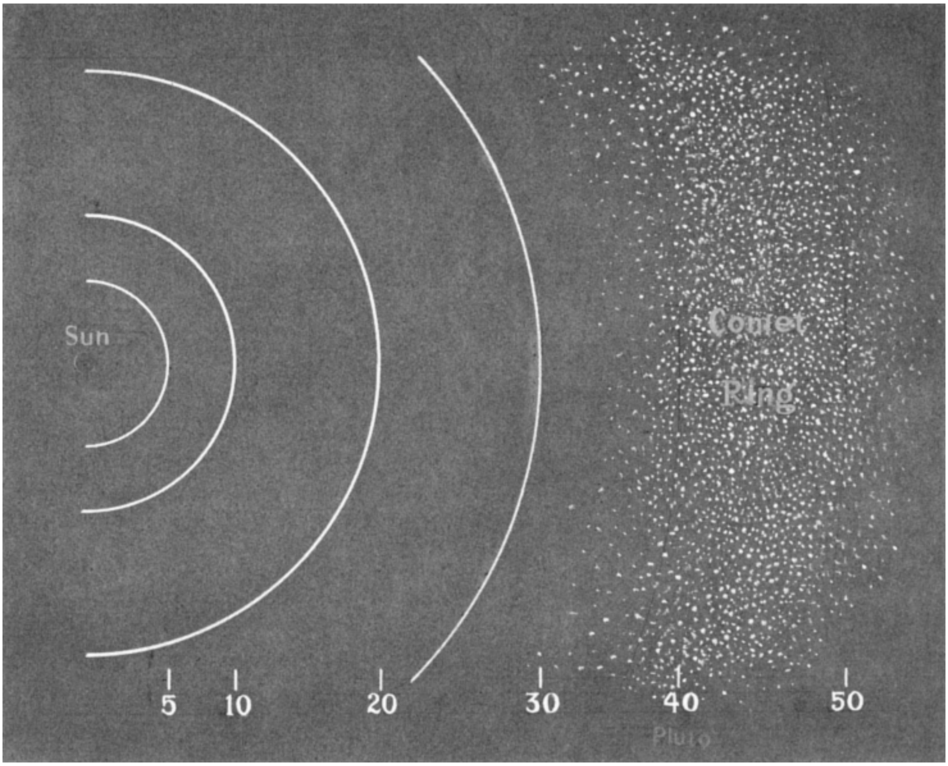


FIGURE 14. Suggested position of the comet belt.

of identical nature, although some of those formed near the disk might contain a smaller abundance of compounds with high vapor pressure at low temperatures, such as methane (CH_4), neon, and some unsaturated molecules.

Let us now examine the evidence for a belt of comets beyond Neptune. Suppose there actually exists a total mass of 10 to 100 times the earth made up of objects 1 km or greater in radius constituting a disk of thickness of some 2° as seen from the earth at a distance of some 40 a.u. If the comets reflect sunlight like the full moon (7 per cent), the disk will have a surface brightness of 7th magnitude per square degree or fainter, less than one-twentieth the brightness of the average night sky. Such a disk would not be apparent above the brightness of the zodiacal light and the Gegenschein, which are a nearby cloud of dust near the plane of the planets seen by scattered sunlight. The comet belt may, however, be a constant contributor to the dust of the zodiacal cloud. Conceivably it might be observed by satellites rising well above the plane of the earth's orbit.

The comet belt, however, would have an appreciable attraction which would disturb the motions of the outer planets. Note that Pluto's mass, so determined, is much too large, even if Pluto be made of gold. The major evidence for Pluto's mass depends upon slight deviations from theory in the latitude of Neptune observed since its discovery in 1846. I find[45] that these deviations can be explained slightly better by a disk of comets than by Pluto. There is no unique solution for the mass and position of the disk. The observations can be fitted, for example, if one assumes that there exists a belt of material at a solar distance of 40 to 50 a.u., tilted 3° to the invariable plane (near Jupiter's orbital plane) and containing 10 to 20 times the earth's mass.

The existence of the belt could probably be established or disproved by the determination of modern, definitive orbits of Neptune, Uranus, and Pluto using the observations now available.

THE FORMATION OF THE TERRESTRIAL PLANETS

Let us return now to the possible sequence of events within Jupiter's orbit. I see no reason to assume that the collapsing protosun should have ceased to leave a Laplacian discoid within Jupiter's present orbit. The density of the discoid per unit area of the plane should have increased fairly uniformly with decreasing solar distance until the protosun reached some temporary equilibrium, perhaps near the dimension of Mercury's present orbit. Any other assumption has a purely *ad hoc* character although many other possibilities cannot be disproved.

All evidence points to an accumulation of earthy planetesimals within Jupiter's orbit.

A combination of three factors, however, could have prevented a large planet from forming within Jupiter's distance. These are :(1) The higher densities and the greater heat of collapse of the discoid within Jupiter's orbit may have kept the ices from freezing while the gases were eliminated from the inner part of the system. (2) The magnetic lines of force were strengthened and the solar activity continuously increased in the later stages of collapse, preventing larger gaseous aggregates from collecting within Jupiter's orbit. (3) Jupiter's mass perturbed the motions of the solid aggregating planetesimals, increasing their orbital inclinations and eccentricities, thus increasing the violence of collisions and decreasing the rate of accumulation.

Before trying to reconstruct the early stages, let us work backwards from the present, at least with regard to the asteroid belt. All our evidence indicates that the asteroid belt is now a dissipative system; that is, collisions are etching away the asteroids, not building them. The velocity of escape even from the largest, Ceres, is only about 0.3 km/sec. while collisional velocities are typically 2 km/sec. or more. Hence, practically all collisions among all asteroids, large or small, cause both participating bodies to lose more mass than they gain. We may conclude without question that all asteroids are now losing mass.

The present rate of dissipation of the asteroid belt, however, is not well established either theoretically or observationally. The etching rate for meteorites in space is too great to permit them to spiral towards the sun by the momentum exchange with solar radiation; i.e., by the Poynting-Robertson effect. We observe the major collisional debris only in the form of fine dust in the earth's neighborhood. The dust particles, if sufficiently small, can spiral in fast enough to avoid collisional destruction. From present rate estimates extrapolated backward in time, we can calculate that the asteroid belt may have lost as little as 1 per cent of its mass over the last four aeons.

Thus, as Sir Harold Jeffreys[46] has stressed, isolated small planetesimals without a gaseous discoid could never have collected into the present asteroids, once Jupiter began perturbing their orbits! No reasonable, though large, increase in the number and total mass of the early planetesimals will improve this situation.

We must conclude, therefore, that the asteroids were formed by one or more of the following classes of processes, which are not mutually exclusive:

(*a*) From cometesimals or dirty snow, the ices and their vapors being lost after solar radiation became appreciable.

(*b*) From planetesimals at a time when there was sufficient gas both to keep the orbits nearly circular near the fundamental plane and to form quasi-permanent atmospheres about the growing asteroids to cushion the collisions.

(*c*) From sizable planetoids that, as in (*a*) or (*b*) above, accumulated rapidly in the early stages and later broke up into relatively small pieces.

The meteorites, whether or not they are typical of the asteroids, provide our only detailed information about these bodies. We cannot yet measure even the mass or density of any asteroid. The variety of mineralogical structures among the meteorites is evident from the fact that the seventeen hundred examples are individuals; almost all are readily distinguishable from the others. They were both formed and metamorphosed under a wide variety of physical circumstances. No simple single process or single sequence of processes can account for them all.

The prevalent chondrules appear to represent rather primitive matter, droplets that were formed and cooled quickly possibly in the early stages of the collapsing nebular disk. It is quite probable that the earthy material was all or mostly vaporized just prior to the major condensation process near the plane of the planets within Jupiter's orbit. The pressures, however, could not have been great enough to have condensed out liquid droplets even at Mercury's orbit. Rather we expect the material, whether earthy or icy, to have condensed initially into rather fluffy, smoke-type particles which then accumulated into larger aggregates. The early ones could have been melted into droplets by heat in the nebula near the sun, very likely by violent shock waves as Wood suggests, or conceivably by falling through the primitive atmospheres of asteroidal bodies. None of these concepts has yet any real theoretical support. The last requires velocities of several kilometers per second, which one might prefer to provide by gravitational attraction; this would then suggest rather massive parent bodies for the meteorites; i.e., the size of the moon or greater. We have seen that slow cooling rates in solid bodies tend to preclude this explanation. Conceivably chondrules may have formed in droplets within pools of molten silicates on the surface of asteroids that still retained their primitive hydrogen atmospheres.

It seems likely, however, that many chondrules were made by the following collisional processes, which certainly operated in all but the very earliest stages of condensation and accumulation. These processes are:

1 *The collisions of "smoke" particles with each other and with larger planetesimals in the early reducing atmosphere* The smoke particles would have been the condensate of the earthy materials with an appreciable amount of adsorbed primitive gas. On collisions at velocities of a very few kilometers per second the smoke particles, having an initial low density, would compress and form droplets which would cool very quickly. Experiments by Henry J. Moore[47] show, indeed, that porous rocks such as volcanic tuff do produce droplets much more efficiently on high-velocity encounter than do high-density particles.

2 *The normal crater-forming collisions among the denser particles* Such collisions always produce an appreciable amount of droplet spray which will cool quickly to form chondrules. The efficiency of this process is lower than for collisions involving low-density smoke particles.

3 *Collisions among massive planetesimals having heated centers* As Urey[48] points out, such collisions will shatter much of the internal structure because of the sudden release of internal pressure involving some residual gas pressure. Such collisions occurring after the central temperatures had reached fairly high values could result in the production of an enormous quantity of small droplets that would cool rapidly; i.e. chondrules. Fredriksson and Ringwood[49] have recently demonstrated the efficacy of this process in the laboratory. The nature of the collision would determine whether the chondrules were released into free space or reassembled gravitationally in a complex reamalgamation of the material contributed by both colliding bodies.

The capture of chondrules on sizable asteroidal bodies could be accomplished with little damage by the cushioning effect of extremely tenuous atmospheres with densities the order of 10^{-9} gm/cc, or less. These atmospheres need not be permanent but could be quasi-permanent, continuously replaced in relatively short intervals of time.

That we find so much meteoritic material with chondritic structure follows immediately from the nature of the collection and dissipation processes for such bodies. Having collected an appreciable amount of the incoming material in the form of chondrules, the asteroid grows, the pressure increases, and the deeper material tends to outgas with the increasing pressure. For asteroids of sufficient size, the eventual radioactive heating produces a molten center in which the iron is segregated gravitationally; but unless the asteroid is quite large there is a considerable volume near the center where the gravitational field is small. There we might expect to find a mixture of iron and stone, the pallasites. Somewhat farther out radially, the melting has become rather complete to produce the achondritic material in which the chondrules have completely lost their identity. At higher levels the pressures and temperatures are modified by the outgassing process and surface cooling to produce the complex structures of chondrites. Near the surface the pressures and temperatures are low so that none of the inflowing material under the relatively low surface gravities melts; it sinters somewhat into material of low structural strength which is easily destroyed by collisions. The same situation applies to very small asteroids, perhaps the order of 30 km in diameter. All such asteroids now in existence would therefore be fragments of larger ones.

For the probably most abundant small asteroids without an iron core the above sequence leads to a final state in which very little gas is retained, even in the upper layers. Most of the asteroid is a chondrite. The nature of the trace of gas depends heavily upon the time at which the asteroid grows and probably also upon its position with respect to the sun. Early formation will occur in a reducing atmosphere. Late formation in larger bodies could result in considerable accumulation of the heavier gases, including water. Also it is possible that some contributions have been made by cometesimals, particularly in the outer regions of the asteroidal belt near Jupiter.

Among asteroids in which the heavier gases are appreciable, the outer cold structure could form an ice trap near the surface, below which water and other volatile materials may have collected — a suggestion by DuFresne and Anders.[50] The carbonaceous chondrites could have developed in this manner.

On collisional breakage of the asteroid, the fragile outer material is quickly lost. If the breakup occurs early, while some of the solar nebula is still present, part of the material may be reaccumulated. Later on it will be ground to fine dust or gas and lost completely to small asteroids. The nickel-iron of the core will persist in space far better than the achondritic and chondritic material. We may today overestimate the original nickel-iron fraction by a large factor (10 or 100 times?). Probably most of the mechanically stronger material was chondritic from the smaller asteroids, the iron cores constituting a very minor fraction of the total asteroidal mass.

In discussing the complex processes of asteroidal formation and disintegration, it is vital to remember that these processes are continuous as stressed in the theory of O. J. Schmidt and B. Levin.[51] At all stages, collisions can produce a type of structure that Urey would class as a *secondary type* of meteorite, distinguished from the *primary* type that was quickly and immediately formed from the solar condensation. Consider two asteroids that collide at low velocities such that the breakage is not dissipative; that is, such that the final mass of the two after encounter is greater than the mass of either one originally. In such encounters an enormous amount of crushing and cracking will occur. This permits a certain redistribution of material between superficial and deeper layers. Immediately after the collision the broken mass will be rounded in form by gravity and the interior sealed.

If such collisions occur in the early stages of the nebula, they will produce bodies that continue through the cycle of radioactive heating so that their centers may be indistinguishable from those that were formed, particle by particle, without large disruptive collisional effects. Collisions of this type at successively later periods will produce meteorites of increasing complexity and variety. The non-dissipative and moderately dissipative collisions can lead to anomalous juxtaposition of minerals with strikingly different histories, both with regard to temperature and composition, and subjected to further conditioning. Thus it becomes apparent that Urey's concept of primary and secondary bodies can be modified to avoid the difficulty of two distinct accumulation periods or types of processes among the planetesimals. We need not wonder at the enormous variation among meteorites.

Non-dissipative collisions would be limited to successively larger bodies as gas from the nebula is eliminated. In the very early stages the relative velocities among the smoke particles and the early accumulations would be small, allowing a rapid accumulation process. Perhaps magnetic attraction is important, as Wood suggests. After Jupiter perturbations have increased the relative velocities of the planetesimals, collisions among the smaller ones with negligible gravitational fields would become increasingly dissipative. Collisional dissipation of lunar-sized planetesimals would appear to be a rare phenomenon at any stage because of the high velocities required to overcome the internal potential energy. Consequently it seems unlikely that any asteroids larger than the moon ever developed and dissipated by collision. A very few of lunar size or greater may have been enveloped by the terrestrial planets and conceivably the moon may have been captured by the earth after it reached approximately its present size.

The detailed mechanism for eliminating the gases from the discoid is not yet clear, except that it must involve magnetohydrodynamics and violent solar activity, visualized as giant solar flares. A marked divergence from Hoyle's picture of planetary evolution, however, is evident. He calculates that the condensation process of the protosun continued without the formation of an appreciable discoid initially beyond a diameter roughly that of Mercury's present orbit. The magnetic lines of force in the sun then strengthened to approximately a field of one gauss near the surface. The angular momentum transfer to reduce the sun's rotation was effected through the loss of matter which was given higher and higher angular momentum in spiraling and twisted lines of force (Figure 6). He calculates that the outgoing gas carried with it cometesimals up to dimensions the order of 10 meters and that these, plus the outgoing gases, condensed to form the various planets as we observe them.

Unfortunately, I cannot accept part of Hoyle's argument. His evidence that outflowing gas under these circumstances should carry along solid particles appears to be wrong. In the situation he presents, the magnetic lines exert a pressure on each other and tend to separate. In his diagram, the major effect on the gas is an outward pressure from the sun and a smaller component directed in the forward sense of rotation due to the lines of force. The outward pressure, as Hoyle indicates, will not remove the gas from the system, the removal being accomplished by the much smaller component of force forward along the direction of motion, which causes the gas to spiral outward and escape. The larger outward pressure, however, partially counteracts the solar attraction, thus effectively causing the gas to move in a smaller gravitational field than that of the solar mass alone. Consequently, at a given distance from the sun the gas will be moving at a lower velocity than the solid particles. The solids effectively meet a resisting medium of gas; this tends to reduce their angular momentum and cause them to spiral slowly toward the sun. Hoyle's process, I believe, fails to eject solid or icy aggregates at distances beyond the initial discoid. Consequently, I must agree with Cameron that some process of momentum transfer by magnetohydrodynamics begins, not within the orbit of Mercury, but at very much greater solar distances.

A major problem concerns the increasing iron content of the planets towards the sun. It seems impossible that the pressure could have been high enough near the plane of the discoid to produce droplets of iron while silicates remained vapor. A remaining process suggested by Urey depends upon the relative strength of iron *versus* silicate particles to withstand collision. Whether the smaller particles tend to spiral inward or outward from the sun does not greatly affect the situation. Rapid spiraling in either direction reduces the rate of capture by the planets. Slow spiraling increases the capture rate. The stony material will more easily be pulverized to extremely fine particles forced out by light pressure or to gas carried out by the solar wind or by solar flares. It is difficult to evaluate the likelihood of this collisional separation process, but one serious objection is outstanding; viz., the process does not operate for the newly condensed smoke but only after the particles are largely differentiated and the gas is dispersed. It hardly seems possible that enough iron could have been added to the planets in the later stages to have produced such a large density variation (refer to Table V).

The iron abundance in the meteorites and planets is possibly high compared to that in the sun. Note that the orbit of Mercury is much more inclined to the mean plane ($4^\circ.5$ to $9^\circ.8$) than those of the other terrestrial planets, much like the sun's equator ($7^\circ.2$). Could it be that Mercury was formed later than the other planets, from a final nebular wisp rotating at an appreciable angle to the original nebula and containing a greater abundance of iron? Such speculation is probably futile but does indicate the wide range of unprovable possibilities.

With respect to the abundances of the elements, the brilliant theory by Fowler, Greenstein, and Hoyle has some observational support. The theory requires heavy particulate irradiation by solar flares of planetesimals to produce: (1) the observed isotopic anomalies among the light elements, (2) the xenon^{129} anomaly, and (3) the production of short-lived radioactive isotopes like Al^{26}, to heat the asteroids quickly. Since FGH require that only 1/10th to 10 per cent of the earth's mass and of the asteroids need be irradiated, the major accumulation could have occurred before the nebula had been largely cleared away. H. Mitler has also shown that enough water could have been supplied in hydrates so that icy cometesimals are not required. Thus we need not postulate such extremely low temperatures near the earth nor meet with the difficulty of removing a large fraction of water from the earth's initial mass. If these ideas are correct, we should perhaps find somewhat higher abundances of the light elements in some of the meteorites than we find on the earth. The absolute values of these abundances are not yet well enough determined to establish or disprove this concept. The cometesimals formed early at great solar distances could contain appreciable quantities of methane and carbon compounds to contribute to the carbonaceous chondrites. Thus their composition may more closely represent the primitive nebula, as is suspected. Until we actually explore individual asteroids by space probes we may never be able to determine the distance from the sun at which any given meteorite originated.

The high and the low iron content groups of chondrites as found by Urey and Craig might conceivably arise from condensation near the sun for the high iron group and at greater distances for the low iron group. No adequate explanation has yet been presented.

Thus, still speaking qualitatively, the asteroids and the terrestrial planets appear to have accumulated from planetesimals. It would appear too that most of the satellites developed much in the same fashion about their primaries. The moon may be an exception.

The data and theory are not yet adequate to permit a choice between the FGH irradiation theory of isotopic distribution and other possible theories depending upon the genesis of the isotopes in stellar interiors, supernovae explosions, and the like. The large number of measurable isotopes should eventually make possible a clear-cut solution of the contribution made to the isotopic abundances by each of a large number of specific processes. The histories of the isotopes should then clarify the history of the terrestrial planets and asteroids.

THE FORMATION OF THE MOON

A. E. Ringwood[52] and D. U. Wise[53] have revitalized George Darwin's earlier concept that the moon developed from the earth. Suess[54] had suggested that rapid rotation might have enabled the earth to lose a primitive atmosphere. Ringwood and Wise attribute the separation to the suddenly increased rotation rate of the earth at the time it melted and the dense iron settled to the core. This reduced the earth's moment of inertia without a change in angular momentum. Since the earth was already in a precarious condition near rotational instability, the increased rate of rotation ejected matter at the equator in a fairly massive ring. From this ring a protomoon formed near the Roche limit, which is quite close to the earth. Tidal friction caused the protomoon to spiral outwards, picking up all the moonlets outside the initial orbit. The particles within the limit were eliminated by a number of processes.

The possibilities of this modern version of Darwin's theory are fascinating indeed; they provide:

1 an explanation for a suddenly increased rotation rate to precipitate a separation;
2 an answer to the very critical problems concerning the loss of the primitive earth atmosphere and excess water;
3 a low density source for the lunar material from the upper earth mantle after partial magmatic separation; and
4 a time delay after the formation of the earth so that short-lived radioactivities might have largely decayed.

The delay includes not only time for the earth to heat and form a core, but also for the ring to coalesce into the moon and for the moon to accumulate many of the outer moonlets. Thus the moon could have formed quite cold and with a much reduced percentage of radioactive heating agents. As a consequence, the moon need not have melted completely, as observations suggest.

A fifth problem may also be solved. MacDonald[55] finds that bodily tidal friction in the earth, extrapolated backwards in time, spirals the moon to the earth's surface in much less than 4.6 aeons. This is added evidence that the moon may have been formed much later than the planets.

Because the moon is unique among satellites in constituting more than 1 per cent of the mass of its primary, many suggestions have been made concerning its capture by the earth. I will not pursue these suggestions here because of their complexity and number.

I hesitantly add another suggestion to the large number already proposed for the origin of the moon. It is based on the greater friability of silicates as compared to metals. If the smaller particles were largely silicates and spiraled either inward or outward from the sun much faster than the irons, a large Saturn-like ring around the earth would collect nearly all the passing fragments. The earth and terrestrial planets, however, would have much smaller capture radii and hence would fail to collect much of the faster spiraling silicate particles. Thus the earth-ring would grow relatively faster than the earth in the later stages of the accumulation period and would collect lower density earthy materials. The ring would finally collect into moonlets and the final sequence would be similar to that described above for the modified Darwin theory. Thus the moon could have developed a lower mean density than the earth and its formation from cold material could have been delayed to reduce the heating problems discussed above. Even so, this alternative solution for the formation of the moon does not dispose of the primitive earth's atmosphere unless we add another assumption.

In conclusion, we have developed some partially supported ideas and are left with a host of unanswered questions concerning the history of the solar system. Our rapidly increasing store of knowledge possibly raises more new questions than it answers. Nevertheless, I believe that many aspects of planetary formation are no longer speculative but have been placed on a sound foundation. We can expect rapid progress in understanding more and more details of the complex processes involved. Direct exploration of the moon, Mars, asteroids, and comets will be invaluable in clarifying the crucial uncertainties now remaining with regard to the formation of the solar system. Our space program should add immeasurably to the historical record of these truly ancient times.

✲

The Origins of the Continents, Oceans, and Atmosphere

H. H. HESS

CHAIRMAN'S NOTE: *Professor H. H. Hess of Princeton University summarized at this point much of what is known or surmised about the early history of the earth, but he has been unable to prepare his paper for publication. I devote these five pages to the earth in order to maintain the continuity of the discussion.*

DR. WHIPPLE estimates that the earth contains only about one five-hundredth of the initial mass of matter from which it was formed. Nearly all the hydrogen, helium, and noble gases, and most of the carbon, nitrogen, oxygen, and other highly volatile substances have been lost.

At some time after its initial formation, the components of the earth separated into a central core and surrounding shells. The core apparently consists of iron and nickel, largely molten but solid at the center. The diameter of the core is about half the diameter of the earth and it contains nearly a third of the total mass. Surrounding the core is a solid shell called the mantle, nearly 3,000 km thick, which is believed to consist principally of iron and magnesium silicates. Outside is a thin crust containing more silicon and oxygen and less iron and magnesium than the mantle, with relatively high concentrations of sodium, potassium, calcium, aluminum, and other non-volatile elements. Over two-thirds of the earth's total mass is in the mantle; the crust contains about four-tenths of 1 per cent.

The oceans, covering 71 per cent of the solid surface, and consisting of oxygen (85 per cent), hydrogen (11 per cent), and a very thin broth of all the other elements (3.5 per cent), make up 0.025 per cent of the earth's mass. The solid and liquid parts of the earth are bathed in a thick layer of gas which becomes rapidly attenuated with height. Near the surface, this gas consists of molecules of nitro-

---

H. H. HESS Princeton University

gen, oxygen, water, and carbon dioxide, together with atoms of argon. At greater heights, it is a plasma of ionized, highly energetic particles, extending out for several earth diameters. The entire gaseous mass is only about a millionth of the mass of the earth.

Like the stars, the earth is "alive" in its interior, but at a very much lower metabolic rate. Whereas the sun emits 2 ergs/gm/sec., the heat flowing out from the interior of the earth is only 4×10^{-8} ergs/gm/sec. For both the stars and the earth, nuclear processes are the principal source of energy. Because of radioactive decay, the rate of heat generation at the beginning of the earth's history as a planet was five to ten times higher than the present rate.

Radioactive decay of thorium and the two isotopes of uranium produces different isotopes of lead. From the ratios of abundance of these radiogenic lead isotopes in surface sediments and rocks, the time of formation of the earth as a solid spheroid, "isolated and complete in itself," can be calculated. This time is about 4.5 billion years ago. The earth is almost as old as the sun, and perhaps a third as old as the galaxy.

The oldest rocks found anywhere on the earth are less than 3 billion years old, one and a half billion years younger than the earth itself. These oldest rocks occur near the centers of the continents. As we go outward toward the edges of the continental platforms, the rocks become progressively younger, being a few hundred million to a few thousand years old in the coastal regions. Thus the continents have grown outward throughout geologic time.

No rocks more than a hundred million years old have yet been found in the ocean basins. This is only 2 per cent of the age of the earth.

The chemical and mineral composition of rocks, and the shapes, sizes, and arrangements of their contained particles, are a detailed record of the conditions under which they were formed. From these properties, and from the fossil remains of organisms, the geologist can read the history of events near the earth's surface in the localities where the rocks are found. But for 35 per cent of the earth's lifetime we have no such record on the continents, and 98 per cent of the record under the sea floor has not yet been found. We are hopeful that by drilling through the thin (less than a few hundred meters) layer of unconsolidated sediments, and through the underlying young volcanic rocks, we may find a buried oceanic record that can carry us further back in time. At present we can only infer, from fragmentary and circumstantial clues, many of the most critical events in the history of our planet, in particular those events that made it possible for life to begin and for biochemical evolution to proceed. In this realm of profound mystery, astronomical, geological, and biological evidence are all equally relevant.

One of the remarkable facts about the earth is the irregularity of its solid surface. As a result, the oceans do not form a continuous shell but are gathered together in deep basins surrounding the great islands called continents. The question of how and when the oceanic basins and the continents were formed is one of the principal problems of earth history. The differences between the two are striking. The crust of lighter silicates is about 35 km thick under the continents, and only about 7 km thick under the oceans. The continental rocks exhibit a very great variety, reflecting the interactions of many geological processes, while the rocks under the deep sea are volcanic, with monotonously similar compositions. The ocean floor is a giant volcano field. Long, narrow, and deep gashes, called trenches, exist in many places around the edges of the ocean basins. Elsewhere the oceanic crust is broken by long cracks or "fracture zones." The crustal plates between these cracks have slipped past each other for distances up to more than a thousand kilometers.

Down the center of the Atlantic Ocean, from Iceland to 55° S latitude, runs a great, curving spine or ridge, 1,500 km wide and one to two thousand meters higher than the deep ocean floor on either side. This ridge curves around Africa into the Indian Ocean. There it joins a ridge which extends eastward below Australia into the eastern Pacific, and thence northward through Easter Island and the Galapagos Islands to the west coast of North America. These ridges are very young compared to the age of the earth, perhaps only a few tens of millions of years old. In some places their axes are marked by long, narrow rift valleys, while elsewhere the central rifts seem to have been recently filled with lava.

From the scarcity of the noble gases on the earth compared to their solar abundances, it seems likely that the proto-earth's atmosphere was lost, and that at one stage in its early history the earth had little or no atmosphere or oceans. The present oceans and atmosphere may have been pressed out of the solid material throughout the earth's lifetime, synchronously with the growth of the continental areas. If so, the ocean basins must have progressively deepened at about the same rates in order to maintain the relative constancy of sea level on the continents. A steady growth of the atmosphere is suggested by the relatively high content of argon 40, a decay product of the radioactive isotope of potassium. Half of this radiogenic argon could have come from the decay of potassium corresponding to that now present in the continental crust. The remainder must have been squeezed out of the mantle. Recent evidence indicates that the potassium content of the outer mantle layers is very low, about equal to that in stony meteorites, and this evidence, combined with measured rates of heat flow from the deep sea floor, strongly suggests that the potassium content of the mantle is relatively uniform through its entire thickness of 2,900 km. Hence a major fraction of the atmospheric argon must have come from great depths. The rates at which it escaped, and thus by inference the rates at which water and other atmospheric gases were squeezed out of the solid earth, are very uncertain.

The probability that radioactive heat is being generated from potassium decay far down in the mantle suggests, however, that convective heat transfer to the surface is occurring; in other words, that the mantle rocks are slowly turning over in giant convection cells, driven by the heat of radioactive decay. Many geophysical arguments have been advanced against this hypothesis of mantle convection, but the geological evidence is hard to circumvent. For example, the heat flow through the ocean floor is very uneven, being high near the axes of the midocean ridges, and low in the neighborhood of the trenches. This uneven distribution could easily be explained if the ridge axes were the loci of the rising limbs of convection cells and the trenches marked the positions of the downward moving limbs. The immense displacements of plates of the oceanic crust along fracture zones could be readily understood if these plates were being dragged along by underlying mantle material, moving in convection currents of different velocities. Recent evidence from paleomagnetism, the history of the earth's magnetic field, has given a new and convincing impetus to the old hypothesis of continental drift. This evidence is strongly supported by the concordant shapes of the continental platforms at the 2,000 meter contour below sea level on the two sides of the Atlantic Ocean. Continental drift can be understood if convection currents occur in the mantle, but without convection the physical obstacles seem insurmountable.

Our failure to discover ancient rocks under the oceans may mean that such rocks do not exist. The ocean floor may have been swept clean, and the oceanic crust continually replaced, by the drag of convection currents in the upper mantle. On this hypothesis, first advanced by Professor Hess, the oceanic crust is continually being formed along the axes of the midocean ridges, the crust is dragged down the sides of the ridges above the horizontal limbs of convection cells in the upper mantle, and across the ocean basins to the great grinding mills of the trenches along the continental margins, where it is broken and dragged downward.

From the standpoint of the origin and early evolution of life, the most important events in the earth's history occurred in the atmosphere. Most geochemists are now convinced that free oxygen was absent in the early atmosphere and that reducing conditions prevailed. Reducing or non-oxidizing conditions were probably essential to allow the build-up of organic molecules and the anaerobic metabolic processes that we believe were the first stages in the development of living things. The first atmospheric constituents may have been methane, ammonia, and water. These are still the constituents of the atmosphere of Jupiter. Alternatively, the carbon may have been in the form of carbon monoxide and carbon dioxide, if not at the beginning, at least at an early stage.

The appearance of free oxygen in our atmosphere is almost certainly a biological phenomenon. It began when plants developed the complex biochemical process of photosynthesis, in which visible sunlight is used to reduce CO_2 and water to carbohydrates, with release of free oxygen. In an earth without an ocean, the organic matter would be oxidized almost as fast as it was produced. Free oxygen, and hence animal life depending on respiration, could not arise. But on our watery planet, organic matter was buried in marine sediments and remained unoxidized. Hence oxygen could accumulate in the atmosphere as well as in the iron oxide of soils and weathered rocks.

✲

The Origins of Life

GEORGE WALD

WE HAVE HAD a century in which to assimilate the concept of organic evolution, but only recently have we begun to understand that this is only part, perhaps the culminating part, of cosmic evolution. We live in a historical universe, one in which stars and galaxies as well as living creatures are born, mature, grow old, and die. That may indeed be true of the universe as a whole; if so, it appears by some recent estimates to be about 20 billion (twenty thousand million) years old. But whatever doubt is held of the transitory nature of the universe, such a Galaxy as ours surely had a beginning, and pursues its course toward an eventual end; and this, the Milky Way, — perhaps 15 billion years old,[1] about 100,000 light years across, and containing about 100 billion stars — provides a quite adequate stage on which to explore the enterprise of life.

Many of our ideas concerning the beginnings of life on this planet have become familiar, since the modern argument was introduced about thirty years ago by J. B. S. Haldane[2] and in much greater detail by A. I. Oparin.[3] I shall review that argument very briefly here, in order to discuss certain aspects of it further, and to examine some of its universal implications.[4]

The earth is about 4.5 billion years old. The first condensations of interstellar gases and dust that formed the earth and other planets are considerably older, probably nearly as old as the sun, perhaps 5 to 6 billion years; but it is about 4.5 billion years since the earth's core separated from its mantle.[5]

It is a curious fact, not yet well understood, that the oldest rocks on the earth wherever found are about 3 billion years old. It may be only so long ago that the earth's crust became reasonably stable. Not long afterward, on a time scale of this magnitude, life must have appeared; for cherts in the Gunflint iron formation in southern Ontario, Canada, which are at least 1.6 billion years old, contain evidences of microscopic organisms, including what seem to be both fungi and

---

GEORGE WALD The Biological Laboratories, Harvard University

algae.[6] If algae, and hence photosynthesis, existed at least 1.6 billion years ago, life in its more primitive forms must go back much further.

How had it begun? We assume that during the early history of the planet, in an atmosphere almost wholly lacking free oxygen, such simple gases as methane, water vapor, ammonia, and hydrogen, later probably also carbon monoxide and dioxide and nitrogen, reacted slowly but continuously with one another to form the smaller organic molecules. Most of this chemistry probably took place in the upper reaches of the atmosphere, activated mainly by ultraviolet radiation from the sun and by electric discharges. Leeched out of the atmosphere over long ages into the waters of the earth, organic molecules accumulated in the seas, and there interacted with one another so that the seas gradually acquired an increasing concentration and variety of such molecules.

Model experiments have shown how such processes can yield amino acids,[7–9] nucleotide bases,[10] and other molecules that enter into the composition of living organisms. One of the most difficult problems is to attempt to understand how such unit structures combined with one another and polymerized against thermodynamic gradients that tended rather toward hydrolysis — how amino acids could have combined to form polypeptides and proteins, and nucleotides to form nucleic acids, apart from the precise activating mechanisms that guide and provide the energy for such syntheses in cells. Several interesting models for amino acid polymerization have been described, yet involving circumstances that seem to differ significantly from those that could have worked on an adequate scale in nature.[11]

It is thought that over long ages such molecules, large and small, collected in the oceans; and there came together eventually to form aggregates which in turn grew more numerous and complex. These competed with one another, so that some aggregates, by virtue of particularly favorable constitution or organization, proved more efficient than others at sweeping organic molecules out of their surroundings, and so grew at the expense of the others — a primitive beginning of natural selection.[3] It is supposed that sometime, somewhere, or perhaps several times in several places, such an aggregate reached a state that an experienced biologist, had one been present, would have been willing to call alive.

A great question concerns the composition of the primitive atmosphere and hence the gases which were available for the synthesis of the unit organic molecules. Urey[12, 13; cf. also 3] has defended strongly the thesis that this was a reducing atmosphere, containing large amounts of methane, hydrogen, ammonia, and water vapor. This is the mixture of gases with which Miller first demonstrated the synthesis of amino acids and other organic molecules.[7] Later Abelson[8] observed amino acid formation in similar experiments with gas mixtures that included carbon monoxide and dioxide, and nitrogen, always in combinations that involved at least one reducing gas. Heyns *et al.*[9] found that adding hydrogen sulfide to similar mixtures led to the synthesis of a variety of organic sulfur compounds.

A central point is that the primitive atmosphere was anaerobic, and all such model experiments are performed in the absence of oxygen. Under those circumstances they lead to the synthesis of organic molecules; were oxygen present, they would end simply in combustions.

The maintenance of a reducing atmosphere depends upon the presence of reasonably high pressures of hydrogen. The lightest of gases, this escapes from the earth's atmosphere so rapidly that it would have maintained a sufficiently high concentration only throughout the planet's early history. Urey has calculated that a hydrogen pressure of 1.5×10^{-3} atm. should suffice to maintain a reducing atmosphere, and that this may have declined to the present level of about 10^{-6} atm. at the earth's surface some 2 billion years ago.[13] Ammonia and methane, which are unstable in the absence of hydrogen, may shortly afterward have left the atmosphere, to be replaced mainly by nitrogen and carbon dioxide. By that time the organic syntheses we have discussed were presumably already well completed.

It should be noted, however, that this type of discussion rests entirely upon thermodynamic considerations, and takes no account of the kinetics of the reactions concerned. It assumes that the great expanses of geological time are sufficient to complete all spontaneous chemical reactions, and to bring all chemical systems into equilibrium. In fact that is by no means true. Abelson,[14] for example, has shown that the spontaneous decarboxylation of alanine requires more than 10 billion years at 20°C to be half-completed; and the same is likely to be true of any first-order reaction with an Arrhenius energy of activation of more than 40,000 cal per mole. Second-order reactions, the rates of which depend, of course, upon concentration as well as activation energy, must in given instances take equally long. Not all possible reactions therefore need to be completed even within geological times; and obviously this must be an important consideration in dealing with the evolution of the atmosphere. It may well be that simple inertia of reaction permitted much more varied mixtures of gases to persist over long periods than could have occurred under equilibrium conditions.

The existence of molecular oxygen in the atmosphere is a separate issue, and so crucial for our argument that however plausible the geochemical reasons for believing that oxygen was absent initially, some further reassurance would not be amiss.[15]

I think that the evolution of cellular metabolism on this planet provides internal evidence that much of the early development of organisms must have proceeded in the absence of oxygen. It would otherwise be difficult to understand the ingenuity that organisms have displayed in developing anaerobic pathways of metabolism. The entire fundamental structure of cellular metabolism is anaerobic; reactions with molecular oxygen appear as a late epiphenomenon, added to an already complete and within its bounds adequate substructure. The nub of the argument is outlined in Table I.

If, as we suppose, life first appeared in an organic medium in the absence of oxygen, it must first have been supported by fermentations[3] — Pasteur's "life

Table I. EVOLUTION OF ENERGY METABOLISM

Anaerobic Phase

(1) *Fermentation:* a chemical source of energy; by-product CO_2

e.g., $C_6H_{12}O_6 \longrightarrow 2\ C_2H_5OH + 2\ CO_2 + 2 \sim P$

(2) *Hexosemonophosphate cycle:* metabolic hydrogen for reductions

$6\ C_6H_{12}O_6 + 6\ H_2O + 12 \sim P \longrightarrow 12\ H_2 + 5\ C_6H_{12}O_6 + 6\ CO_2$

(3) *Photophosphorylation:* light into high-energy phosphates

$$\xrightarrow[\text{Chlorophylls, cytochromes}]{\text{Light}} \sim P$$

(4) *Photosynthesis:* light into new organic molecules; by-product O_2

$$\text{Bacteria: } 6\ CO_2 + 12\ H_2A \xrightarrow[\text{chlorophyll}]{\text{light}} C_6H_{12}O_6 + 6\ H_2A + 12\ A$$

Algae, higher plants:

$$6\ CO_2 + 12\ H_2O \xrightarrow[\text{chlorophyll}]{\text{light}} C_6H_{12}O_6 + 6\ H_2O + 6\ O_2$$

Aerobic Phase

(5) *Respiration:* metabolic energy from combustions

$C_6H_{12}O_6 + 6\ H_2O + 6\ O_2 \longrightarrow 6\ CO_2 + 12\ H_2O + 30\text{–}40 \sim P$

without air." In so far we beg the question. But fermentation remains in a sense the basic way of life. Fermentative processes underlie all other forms of metabolism; and virtually all types of cell can survive for periods on fermentation if deprived of oxygen. Fermentation degrades organic molecules anaerobically, making the free energy so released available to the cell in the form of high-energy phosphates (e.g., adenosine triphosphate, ATP; represented here by the symbol $\sim$P). Certain familiar forms of fermentation, as for instance the alcohol fermentation shown in Table I, produce as by-product carbon dioxide. This gas, like oxygen, was probably in very low concentration in the primitive atmosphere, and its production by fermentation probably played a great role in the further evolution of metabolism.

There is good reason to believe that the next type of metabolism to develop was the hexosemonophosphate (HMP) cycle. Since this has been worked out relatively recently, it still tends to be thought of as rather esoteric; in fact it is basic. It is often characterized as an alternative ("shunt") pathway of respiration; but it is that only when frustrated. In fact it is essentially an anaerobic process, more closely related therefore to fermentation. This process develops hydrogen for organic reductions and reductive syntheses anaerobically, from sugar with the aid of energy derived from ATP. Incidentally it yields carbon dioxide as by-product.[16] It presents us also with a first example of a reaction of fundamental importance, the metabolic splitting of water; for half the hydrogen produced by the HMP cycle is derived ultimately from water.

If the primitive atmosphere contained much hydrogen, it may be asked why cells could not have used this directly, rather than having to produce hydrogen metabolically. The answer is probably, as already indicated, that by the time living cells had developed to this point, all but the last remnants of hydrogen had already escaped from the atmosphere.

The next process to develop was probably photophosphorylation — the direct utilization of sunlight to produce ATP.[17] This involves also the first appearance in metabolism of the metalloporphyrins: the pigment, chlorophyll, a magnesium porphyrin, to absorb the light; and cytochromes, iron porphyrin proteins, to aid in the transduction of the absorbed energy to ATP.

With that the way was open to a fourth development, photosynthesis, largely an integration of developments already achieved in steps (2) and (3). In photosynthesis the energy of sunlight, transduced through chlorophylls and then in part through ATP, is used to synthesize glucose on the basis of a modified HMP cycle running in reverse.[16, 17] The over-all process involves the splitting of hydrogen from a donor molecule, and its use to reduce carbon dioxide to carbohydrate. A variety of organic and inorganic molecules serve as hydrogen donors in photosynthetic bacteria; but in algae and higher green plants water itself donates the hydrogen, and molecular oxygen is released as by-product.

This is the means by which molecular oxygen entered our atmosphere. When it had reached a sufficient concentration — about 10^{-3} to 10^{-2} atmospheres at sea level — that at last made possible the first aerobic form of metabolism, cellular respiration. In its over-all effects, and even to a large extent in its mechanisms, respiration is the reverse of algal and higher plant photosynthesis. Chemical energy obtained by the combustion of glucose and other organic molecules is made available in the form of ATP, with carbon dioxide and water as principal end-products.

Since its advent, respiration and the reverse process of photosynthesis have been pitted against each other. Presumably they came into balance ages ago. Yet there must have been a great interval in which organisms were slowly turning an anaerobic into an aerobic world; and the increase of oxygen in our atmosphere from negligible beginnings to its present content of 21 per cent testifies to the long period in which photosynthesis overbalanced respiration.

It is usual to think of the physical environment as given, as the absolute setting to which organisms must at all times adapt if they are to survive. It is becoming plain, however, that some of the salient features of our physical environment are themselves the work of living organisms. They not only put molecular oxygen into the atmosphere. By now also organisms have spread upon the earth on such a scale that the atmosphere and hydrosphere have become components in their metabolism. It is estimated that at present all the oxygen in the atmosphere passes through organisms — in by respiration and out by photosynthesis — every 2 thousand years; that all the carbon dioxide in both the atmosphere and hydrosphere cycles through organisms in the reverse direction every 3 hundred years; and that all the waters of the earth are decomposed and recomposed by photosynthesis and respiration every 2 million years.[18]

The combustion of organic molecules is an over-all effect of respiration, but not its mechanism. The actual mechanism is peculiarly significant for our problem. Biological oxidations, with rare exceptions, are performed, not by adding oxygen, but by removing hydrogen. Even when the oxygen content of some molecule in the organism is increased, this is almost always done by adding water and removing hydrogen. Organisms are remarkably adept at performing their oxidations anaerobically. Their only direct combustion is the burning of hydrogen; and the incorporation of part of the energy of this process into ATP is the principal contribution of cellular respiration. But this is still, with relatively few exceptions, the only use that organisms make of molecular oxygen.

The point can perhaps be made clearer with a simple industrial analogy (Table II). Coal can be burned in either of two ways: directly with oxygen to

Table II. ALTERNATIVE WAYS OF BURNING CARBON

Aerobic

$C + O_2 \longrightarrow CO_2$ (combustion of coal)

Anaerobic—Aerobic

$C + H_2O \longrightarrow CO + H_2$ (water gas)

$CO + H_2O \longrightarrow CO_2 + H_2$ (industrial production of hydrogen)

$2\,H_2 + O_2 \longrightarrow 2\,H_2O$ (combustion of hydrogen)

Metabolism of Living Organisms

$(CH_2O)_n$ = carbohydrate (*n* equivalents of water gas)

$(CH_2O) + H_2O \longrightarrow CO_2 + H_2$ (glycolysis)

$2\,H_2 + O_2 \longrightarrow 2\,H_2O$ (respiration)

carbon dioxide, as in a furnace; or it can be used to draw an atom of oxygen out of water, yielding as products the inflammable mixture of carbon monoxide and hydrogen called water gas. Water gas in turn might simply be burned with oxygen to carbon dioxide and water; but instead the carbon monoxide can be used to draw another atom of oxygen out of water, yielding another hydrogen molecule, as in the process for the industrial production of hydrogen. Finally the hydrogen can be burned with oxygen to water.[19]

Glucose, the principal metabolite for energy production in living organisms, is a form of carbohydrate, the unit structure of which is CH_2O (taken six times over in glucose and other hexoses, $C_6H_{12}O_6$). This has the same composition as water gas. Like water gas, a unit of carbohydrate could be burned with one molecule of oxygen to CO_2 and H_2O; but that is not the organism's way. Instead, the carbohydrate is used to split water according to the fundamental equation, $CH_2O + H_2O \rightarrow CO_2 + 2\,H_2$, a somewhat disguised equation of preparatory glycolysis, exactly analogous to the industrial production of hydrogen. This is then followed by the combustion of hydrogen, $2\ H_2 + O_2 \rightarrow 2\ H_2O$, the fundamental equation of cellular respiration.

Horowitz[20] has proposed a plausible mechanism by which biosynthetic pathways may have evolved during the period when organic metabolites were still plentiful in the environment. He suggests that biosynthetic sequences developed in reverse, starting at the end of the chain, and working backward by adding one enzyme at a time. An organism dependent upon some available metabolite, as that became depleted might develop an enzyme for performing the last step in its synthesis, so that the organism's needs were now transferred to the immediate precursor. As that in turn became depleted, the organism might evolve an enzyme for the previous step in synthesis. So it would work its way step by step backward through the sequence until the entire synthesis could be performed from simple and readily obtainable precursors.

This seems then to have been the history of life upon this planet: the slow combination of the gases of the primitive atmosphere to form simple unit organic molecules which accumulated in the sea; the polymerization of some of those molecules to form the first macromolecules comparable with our present proteins and nucleic acids; the aggregation of such large and small molecules in the sea to form micelles of various sizes and grades of complexity, with the final achievement of the living state. Then the gradual mastery of the fundamental problems of deriving energy and preparing new organic molecules with which life could eventually spread upon a cosmic scale, in the process transforming radically the atmosphere of the planet.

I think that some such account as this would now be widely accepted as describing the origin of life on the earth. What is perhaps more interesting is the dawning realization that this problem involves universal elements, that life in fact is probably a universal phenomenon, bound to occur wherever in the universe conditions permit and sufficient time has elapsed.

Those conditions almost surely involve a planet somewhat resembling the earth: of about this size and temperature, and receiving about this quality and amount of radiation from its sun. To mention a few points of the argument: a much smaller planet could not hold an adequate atmosphere, a much larger one might hold too dense an atmosphere to permit radiation to penetrate to its surface. Too cold a planet would slow down too greatly the chemical reactions by which life arises; too warm a planet would be incompatible with the orderly existence of macromolecules. The limits of temperature are probably close to those at which water remains liquid, itself almost surely a necessary condition for life. Life can arise without continuously absorbing radiation, though as we have noted radiation prepares the way by activating organic syntheses in the atmosphere; but it is difficult to see how life can go far, or even persist indefinitely without an external source of energy. By now all life upon the earth runs on sunlight, with the exception of a few chemosynthetic bacteria. Not all radiations are adequate; a range of wave lengths between about 300 and 1100 mμ is needed. Shorter wave lengths than 300 mμ destroy macromolecules; on the earth they denature proteins and depolymerize nucleic acids. Longer wave lengths than about 1100 mμ involve quanta too small to excite molecules electronically, and hence to activate photochemical reactions.

How many such planets exist? By present estimates about 1 to 5 per cent of the stars in our galaxy might possess planets capable of supporting life.[22] That would mean at least 1 billion such planets in our galaxy alone; and since there are about 100 million galaxies now within range of the most powerful telescopes, the number of planets suitable for life in the already observed universe may be of the order of 10^{17}. This number is so vast — even if it were reduced a million times — as to make it difficult to avoid the conclusion that life is widespread in the universe.

On this planet, living organisms are composed almost entirely of 16 to 21 elements — 16 found in almost all organisms, 5 more restricted to particular groups (Table III). A first striking regularity is that these tend to be light elements. All the bioelements except molybdenum and iodine occur within the lightest 30 of the 92 natural elements. That in itself does not seem strange, for the

Table III. An Alphabet of Organisms

Elementary particles	Bioelements	Unit molecules
	Water: H, O	
Protons	*Organic:* H, C, N, O; P, S	Glucose, ribose
Electrons	*Ions:* Na^+, K^+, Mg^{++}, Ca^{++}, Cl^-	Fat, phosphatide
Neutrons	*Trace elements:* Mn, Fe, Co, Cu,	20 amino acids
Photons	Zn (B, Al, V, Mo, I)*	5 nucleotide bases
4	16–21	29

* The elements within parentheses are restricted to special groups of organisms; the others are very generally distributed.

lightest elements tend also to be the most abundant, on the earth as elsewhere in the cosmos.

The bioelements fall into three natural groups, according to the uses organisms make of them: those that form water and the bulk of the organic molecules; the monatomic ions; and the trace elements.

The group of monatomic ions may have been chosen mainly on the basis of their relative abundance, though I do not think that is the whole story even with them. They are the ions most prevalent in the sea; and that may account largely not only for their appearance in cells, but for the fact that animal blood tends to resemble sea water so closely. Most marine invertebrates circulate a solution of ions in their bloods that is essentially sea water. The ionic composition of vertebrate blood so closely resembles sea water diluted three to four times as to have prompted the suggestion that it represents sea water of the remote period in which the ancestors of the vertebrates closed off their circulations.[23]

An argument from abundance, however, cannot be consistently maintained for the other two groups of bioelements. Some of them are abundant; others of equal importance are relatively rare, and organisms concentrate them many times over out of very dilute environments. These elements were selected on the basis of their essential properties rather than their availability; and that is true to a degree even of the monatomic ions. One has a strong indication of this in the fact that the three groups into which the bioelements are divided functionally are reflected in their positions in the Periodic System, where, except for the halogens Cl and I, they form three clusters: the organic elements at the right in the first three periods, the monatomic cations at the left of the third and fourth periods, the major trace elements toward the middle of the fourth (the first long) period, all of them but zinc being so-called transition elements. Being transition elements, these last readily form complexes, as does also zinc; and — being atoms of variable valence — they readily exchange electrons, properties that fit them admirably for the roles we find them principally playing in organisms, as the nuclei of metallo-organic complexes and oxido-reduction enzymes.

The most important such argument, however, involves the four elements hydrogen, oxygen, nitrogen, and carbon, which together make up about 99 per cent of the living parts of living organisms. I think that a responsible examination of the possibilities leads inevitably to the conclusion that life everywhere in the universe must be constructed primarily of these four elements.

Hydrogen, of course, is the most abundant element in the universe, and hydrogen and oxygen constitute large fractions (15.40 and 55.19 per cent) of the atoms in those portions of the earth accessible to living organisms (the whole hydrosphere and atmosphere, and the crust to a depth of 10 miles).[24] On the other hand carbon must be extracted from the last 0.44 per cent of the accessible atoms, and nitrogen from the last 0.16 per cent (one may add that phosphorus must be found in the last 0.23 per cent and sulfur in the last 0.12 per cent of the accessible atoms). These are therefore not consistently the most plentiful elements. They owe their status rather to their essential "fitness": they alone among the natural elements possess the critical properties upon which the existence of life everywhere in the universe must depend.

(I have been asked sometimes how one can be sure that elsewhere in the universe there may not be further elements, other than those in the Periodic System. I have tried to answer by saying, it is like asking how one knows that elsewhere in the universe there may not be another whole number between 4 and 5. Unfortunately, some persons think that is a good question too.)

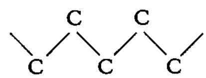
We are — understandably — so greatly impressed with the regularities in the Periodic System as to sometimes exaggerate them. The lightest elements, specifically those in the first two periods, in fact exhibit quite distinctive properties, not repeated in the lower periods. It hardly needs urging that silicon has quite different properties from carbon, phosphorus from nitrogen, and sulfur from oxygen. Hydrogen, of course, has wholly unique properties.

The special distinction of hydrogen, oxygen, nitrogen, and carbon is that they are the four smallest elements in the Periodic System that achieve stable electronic configurations by gaining respectively 1, 2, 3, and 4 electrons. Gaining electrons, in the form of sharing them with other atoms, is the means of making chemical bonds, and so of making molecules. The special point of smallness is that these smallest elements make the tightest bonds and so the most stable molecules; and that carbon, nitrogen, and oxygen are the only elements that regularly form double and triple bonds. Both properties are critically important.

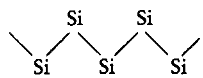
It is frequently suggested that elsewhere in the universe silicon may substitute for carbon in living organisms. The reasons for considering silicon are that it falls just below carbon in the Periodic System; like carbon it can combine with itself to form long chains, and hence very complex molecules; and here on earth there is about 135 times as much silicon as carbon in the areas accessible to life.

Silicon, however, cannot replace carbon in living organisms.[25] For one thing, it forms looser, less stable compounds, but that, though a disadvantage, might be tolerated (Table IV). Silicon chains, however, are susceptible to attack by molecules possessing lone pairs of electrons, in part because of their open structure, but still more because silicon, a third-period element, possesses *3d* orbitals

Table IV. CARBON AND SILICON CHAINS



C – C distance, 1.54 A.
bond energy, 83.1 kcal per mole



Si – Si distance, 2.34 A.
bond energy, 42.5 kcal per mole

Such silicon chains are unstable to O_2, NH_3, H_2O, lone electron pairs of which attack by occupying *3d* orbitals of Si.

available for further combination. For this reason silicon chains cannot exist for long in the presence of oxygen, ammonia, or water. I think that in itself must eliminate silicon as a possible basis for life.

Silicon, however, has another fatal disability, its failure to form multiple bonds. The importance of this factor can be understood if one compares carbon dioxide with silicon dioxide (Table V). In carbon dioxide, double bonds between the carbon and oxygen atoms completely saturate their combining capacities. The molecule goes off freely as a gas, and dissolves in and combines with water, the forms in which organisms obtain it. In silicon dioxide, however, silicon remains singly-bonded to oxygen, leaving four unpaired electrons on the molecule. These promptly form bonds with adjacent silicon dioxide molecules, and they in turn with others. The result is a huge polymer, a supermolecule such as quartz, so hard because it can be broken only by breaking covalent bonds. That is why silicon is fit for making quartz, but living organisms must be made of carbon.

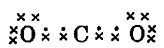
Somewhat less compelling arguments of special fitness involve phosphorus and sulfur,[25] which among the other functions they perform in organisms, have the special role of forming high-energy compounds (e.g., ATP, acyl coenzyme

A) which transfer energy and organic groups within the cell. The very factors that constitute a disability in silicon become an advantage in phosphorus and sulfur (Table VI). The openness of their bonds and their possession of 3*d* orbitals capable of receiving further electrons make their compounds susceptible to attack by molecules that can offer lone pairs of electrons; and this provides a mechanism for the participation of phosphorus and sulfur compounds in the energy- and group-transfer reactions that constitute their principal contribution to cellular metabolism.

The major bioelements therefore present unique properties indispensable for the formation and function of living organisms. They — particularly carbon, hydrogen, nitrogen, and oxygen — form also a number of unique molecules, indispensable or of quite singular importance for organisms. Of these the chief is

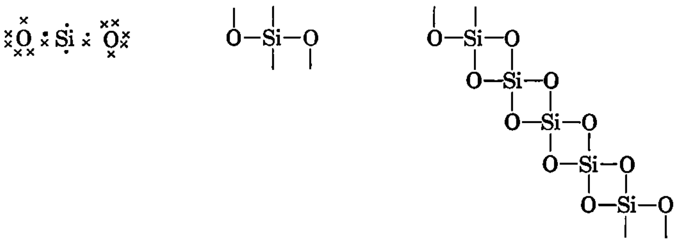
Table V. CARBON DIOXIDE AND SILICON DIOXIDE

Carbon dioxide: CO_2



O=C=O

Silicon dioxide: $(SiO_2)_n$



(The diagram of the silicon dioxide polymer is intended only to represent the mutual saturation of valences, not at all the spatial relationships in the crystal.)

Table VI. "HIGH-ENERGY BONDS" OF P AND S

C—O—P (1.431, 1.76) P—O—P (1.76, 1.76) N—P (1.80) C—S (1.81)	The valence electrons of P and S are in the third shell, which, beyond holding 8 electrons in its *s* and *p* orbitals, can accept further electrons in its *d* orbitals. This, in addition to the large bond radii of P and S, opens the compounds of these elements to attack by molecules possessing lone pairs of electrons, such as H_2O.

water, which I believe to be altogether indispensable. Carbon dioxide must be hardly less important, a gas highly soluble in water, which therefore permeates the atmosphere and hydrosphere, and so is uniquely suitable for circulating carbon among organisms. Carbon dioxide possesses many other fortunate properties — great stability, high density, the capacity rapidly to achieve complex equilibria involving the gas, carbonic acid, and solid carbonates and bicarbonates — all potentially of the highest importance for the formation and maintenance of life.[26]

As already noted, organisms seem to have arisen on Earth, and might conceivably have persisted indefinitely in the absence of molecular oxygen. Yet oxygen, in permitting the development of cellular respiration, provides by far the most efficient chemical source of energy available to organisms, based upon the most energetic of combustions, that of hydrogen. On this planet organisms, when not living directly upon the energy of sunlight, live for the most part on respiration. Living organisms everywhere should have to solve sooner or later the problem of obtaining energy economically by chemical means, in order to survive periods of darkness, even if only cycles of day and night. It is doubtful that such needs can be met more effectively than by combustions. Oxygen therefore, though not indispensable for life, must be reckoned among the special molecules needed for its fullest development.

It has often been suggested that elsewhere in the universe liquid ammonia may take the place of water as a substrate for life. In most important regards, however, liquid ammonia is inferior to water in the properties upon which organisms most depend.[26] An obvious disadvantage is that ammonia remains liquid over a much narrower temperature range than water (Table VII); keeping

Table VII. Properties of Ammonia

Boiling point: −33.35°C (at 1 atmosphere pressure)
Freezing point: −77.7° (at 1 atmosphere pressure)
Density at −79°C
Liquid: 0.7354 gm per cm^3*
Solid: 0.817 gm per cm^3‡

* Cragoe, C. S., and Harper, D. R., 3rd, *Scientific Papers, United States Bureau of Standards,* 1921, **17**, 287.
‡ McKelvy, E. C., and Taylor, C. S., *Scientific Papers, United States Bureau of Standards,* 1922–23, **18**, 655.

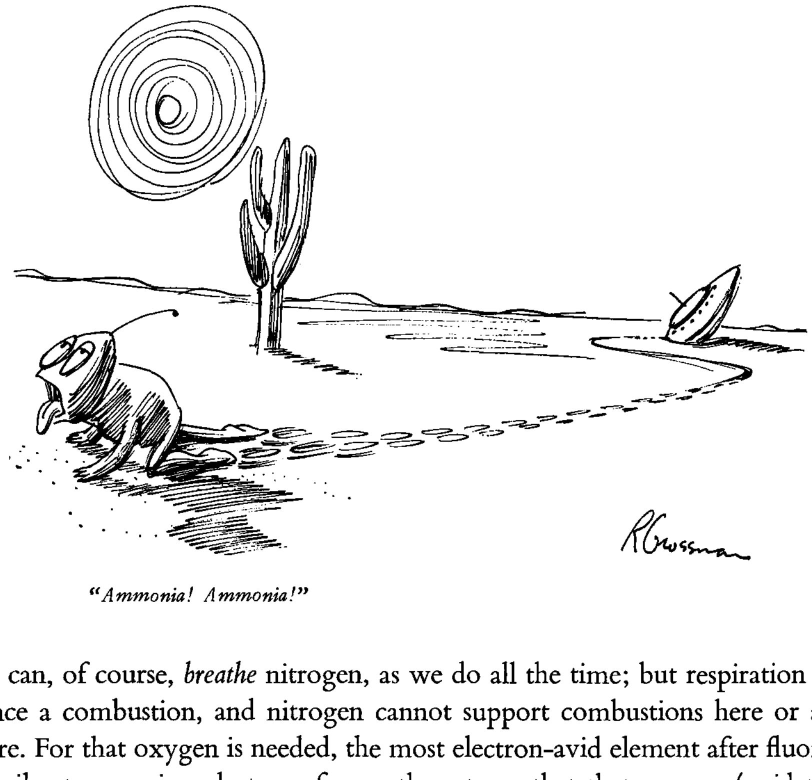
ammonia liquid therefore demands a greater stability of temperature over long periods than many planets afford. But to keep it liquid demands also temperatures that would never rise above perhaps $-40°C$, that is at least 60° below prevailing temperatures upon the earth. (Raising the atmospheric pressure considerably would, of course, permit ammonia to stay liquid at higher temperatures, but would introduce other, equally serious difficulties.) If we assume the modest temperature coefficient ($Q_{10°}$) of 2, chemical processes in general would take about 2^6 as long to complete at $-40°C$ as upon the earth. The processes that led to the origin of life within perhaps a billion years upon this planet might then take some 64 billion years in an environment of liquid ammonia. That is far greater than any present estimates of the age of our Galaxy, and much longer than a star like the sun could remain on the main sequence — i.e., remain adequately supplied with hydrogen to maintain conditions that would make life possible on one of its planets.

L. J. Henderson,[26] in his classic exposition of the many extraordinary properties which make water critically important for living organisms, made much of one of its most peculiar properties — that ice floats. If ice did not float, the waters of the earth would long ago have frozen solid. Even relatively long periods of warm weather could not have thawed them; and life arising or persisting under such conditions would be well-nigh impossible.

Water owes this strange property to the fact that though on cooling it contracts, as do other well behaved substances, below 4°C it expands, so that at its freezing point, 0°, it has a lower density than liquid water. The reason for this is that below 4°C water molecules become increasingly hydrogen-bonded to one another. By the freezing point every hydrogen atom is engaged in both covalent and hydrogen bonding to adjacent oxygen atoms, holding all the water molecules rigidly in an open structure in which they are less densely packed than in liquid water.

Some years ago I began to wonder whether ammonia ice floats. Unable to find this information in the literature, I was glad to have the question answered for me experimentally.[27] Later, wanting to be doubly sure, I repeated the experiment myself; and just afterward found that the information had been published 20 years before (Table VII). Ammonia ice sinks in liquid ammonia, rapidly and unequivocally, hitting the bottom of the vessel with a distinct thud. That can be added to the disabilities of ammonia as a medium for life.

The splitting of water in higher plant photosynthesis yields as by-product molecular oxygen, permitting eventually the development of cellular respiration. A comparable process performed with ammonia might yield instead nitrogen, a depressingly inert gas. An English astronomer was reported recently to have said that there may be places in the universe where life is based on ammonia rather than water, and where living creatures respire nitrogen instead of oxygen. This, among other effects, inspired the accompanying work of art.



"Ammonia! Ammonia!"

One can, of course, *breathe* nitrogen, as we do all the time; but respiration is in essence a combustion, and nitrogen cannot support combustions here or anywhere. For that oxygen is needed, the most electron-avid element after fluorine, so facile at removing electrons from other atoms that that process (oxidation) has been named for it.

For these and similar reasons I have become convinced that life everywhere must be based primarily upon carbon, hydrogen, nitrogen, and oxygen, upon an organic chemistry therefore much as on the earth; and that it can arise only in an environment rich in water. Though the preparatory geochemical syntheses of organic molecules seem to demand an anaerobic environment, and are probably fostered by a generally reducing atmosphere, the later history of these developments must be greatly furthered by the more or less automatic appearance of carbon dioxide as the atmosphere loses hydrogen; and living organisms having arisen would in time almost surely find their way to the production and utilization of molecular oxygen.

How formidable a condition is the geochemical accumulation of unit organic molecules — the building blocks of which macromolecules and eventually the first primitive organisms are to be composed? How many such units are needed, at a minimum? I think perhaps fewer than is generally supposed.

I have had the experience lately of introducing young students, many of whom had not studied chemistry before, to some of the basic essentials of biochemistry. I build the subject up from the elementary particles, then the atoms we have been discussing, to end with what I call an alphabet of biochemistry (Table III). It turns out that about 29 organic molecules are enough to introduce the bare essentials. They include glucose, the major product of photosynthesis and major source of metabolic energy and hydrogen; fats as a principal storage form of metabolic energy; phosphatides as a means of circulating lipids in aqueous media and for their remarkable structure-forming proclivities. Then the 20 amino acids from which all proteins, including all enzymes, are derived. Five nitrogenous bases (adenine, guanine, cytosine, uracil, thymine), together with ribose or its simple derivative deoxyribose and phosphoric acid, form all the nucleic acids, both RNA and DNA. These 29 molecules give students a first entry into the structures of proteins and nucleic acids, the coding of genetic information, the structures of enzymes, the composition and general properties of cell structures, and bring them to a point from which they can begin to explore the complexities of energy metabolism. That this is not the whole of biochemistry goes without saying; the extraordinary thing is that it makes so good a start. Yet this alphabet of biochemistry is hardly longer than our verbal alphabet. That seems to me to imply that the provision of unit molecules preparatory to the rise of living organisms is a reasonably limited enterprise. I cannot help but feel that this situation, that makes it relatively easy to give young students a first taste of biochemistry, must have made it easier also for the first primitive cells to obtain the molecules they needed.

Many of these molecules display in solution the property of optical activity, the capacity to rotate the plane of polarized light to the right or left. This property is almost uniquely associated with the components and products of living organisms; it is as characteristic of life as any property we know. It has its source in so-called asymmetric carbon atoms, carbon atoms bonded to four different groups. All molecules possessing such asymmetric atoms can exist in right- and left-handed forms (dextro- and levo-, D- and L-). When synthesized artificially, such molecules always emerge as equal mixtures of the D- and L-forms, hence optically inactive. Living organisms, however, invariably incorporate one form or the other. Thus virtually all natural amino acids are L-, all natural sugars D-, α-phosphatides L-, and so on.[28]

The point is not that L-amino acids are intrinsically better suited for living organisms than D-, or D-sugars than L-, but that organisms derive important advantages individually and collectively from working consistently with one configuration or the other.[29] Large portions of native proteins exist in the form of the α-helix. They assume that configuration spontaneously; but could do so only with great difficulty if at all, were they made of mixtures of both D- and L-amino acids. This is an important consideration, since most of the specific properties of proteins depend in part upon this feature of their geometry. Similarly the two-stranded helical structure characteristic of DNA and of long sections of RNA demands specific choices in the configurations of three asymmetric carbon atoms in deoxyribose or ribose — carbon atoms 1, 3, and 4. Enzymes, being proteins, are themselves optically active, and in many cases react only with L- or D-substrates, not with both. For these and other reasons organisms constantly choose one configuration or the other, though in each category of molecule either choice would do equally well. Since also the molecules of organisms are in constant flux and interchange, and are passed about widely from one organism to another in complex food chains, there is enormous advantage in staying with consistent series of configurations throughout the whole metabolism, and indeed throughout the population of the planet. For this reason I would suppose that biota that incorporate amino acids on other planets divide about equally between the L- and D-configurations, keeping their other choices consistent with this one.

To go a step further, I think that when confronted with the necessity to develop a molecule to perform some basic cellular function, organisms are highly limited in their choices, though not so limited in their first choices as in their last. Such molecules as the chlorophylls for photosynthesis, the heme pigments for cellular respiration, the carotenoids and vitamins A for photoreception, all represent the outcome on this planet of long and rigorous selective processes that tended constantly toward achieving optimal solutions. All these molecules possess properties that fit them particularly to perform their functions in organisms; and I have no doubt that the better we come in each case to understand the nature of the problem, the clearer it will be why those molecules and not others were selected. Sometimes these molecules present strange mixtures of fortunate and disadvantageous properties; the chlorophylls for example all have the strange property of absorbing light most poorly at those wave lengths at which sunlight at the surface of the earth or under water is most intense. Obviously the chlorophylls must possess other advantages for photosynthesis that far outweigh this disability; and those advantages, since they have given the chlorophylls a unique status in photosynthesis on the earth, might be equally effective in promoting their selection elsewhere.[21]

Again, three animal phyla on the earth, having developed three very different kinds of eye in complete independence from one another, have all arrived at the same molecule, 11-*cis* retinene (11-*cis* vitamin A aldehyde) as the chromophore of their visual pigments. Yet the 11-*cis* isomer is an improbable, intrinsically unstable variant of retinene. Why then choose it repeatedly for this function? It has turned out that the only action of light in vision is to change the shape of a molecule; and 11-*cis* retinene fills that role in an exemplary way, light isomerizing it from the bent and twisted 11-*cis* configuration to the relatively straight all-*trans* with high efficiency. The same forces that guided the selection of this improbable molecule three times independently on this planet might well arrive at the same or similar solutions elsewhere.[21, 30]

To sum up, faced with well-nigh universal problems organisms everywhere may tend to gravitate toward common solutions, types of molecule that within the bounds of organic structure may represent optimal or near optimal solutions. I say types rather than individual molecules, since in each of the cases mentioned we find here upon the earth not one molecule but a category of them at work: at least five different chlorophylls, a variety of hemes, several active carotenoids, two vitamins A.

Such choices must be governed everywhere by natural selection, the process described a century ago by Darwin and Wallace. This is at once the formative and creative principle in the evolution of living things. It involves three components: a mechanism of inheritance, without which life could not continue to exist anywhere; a continuous intrusion of "noise" into the genetic message, appearing in the offspring as random inherited variations (mutations); and the struggle for existence, the competition for the necessities of life, any temporary alleviation of which is met with a leap in population that brings it back into force. These are universal elements, hardly to be avoided in any population of living things. Their outcome is the survival of the fittest — the continuous trend toward optimization, the effects of which on molecular design were invoked above.

It has sometimes been argued that natural selection is "not enough" — not enough to account for the evolution of an eye, or a wing, or the near perfection of embryonic development, or the mating behavior of gulls. But one cannot dismiss natural selection just because it works better than one thinks it should. A hypothesis should be damned for its failures, not its successes—for cases in which evolution has appeared to work to the net disadvantage of organisms. That is, of course, just the problem raised by some instances of the extinction of species in the course of evolution; and some cases of extinction do represent a failure of natural selection, owing we believe to the inertia of the selective process, which on occasion operates too slowly to cope with abrupt changes in the conditions of existence.

Wigner[31] has recently remarked upon "what appears to be a miracle from the point of view of the physicist: that there are (living) structures which produce further identical structures." Fortunately no such miracle occurs. If it did, heredity might seem to work better, but natural selection would not work at all. Wigner offers a calculation to show the quantum-mechanical impossibility of keeping the information coded in the genes from growing increasingly disordered as it is transmitted. The point is that the genetic message is continuously disordered by mutation; but that the selective process as continuously prunes it back to orderly, and indeed toward optimal, sequences. Wigner's calculation can be turned to positive account; it provides some assurance that any molecular genetic code must continually produce such random variations as natural selection demands. Order in living organisms is introduced not beforehand, by preconceived design, but after the fact — the fact of random mutation — by a process akin to editing. We are the products of editing rather than of authorship.

I have thought sometimes that it would be interesting to set up an experimental model for some features of this process. Suppose that random words of the English language were fed out on a tape, and that half a dozen persons independently selected words out of the same random sequence so as to compose first a meaningful sentence, from this beginning a paragraph, and eventually a connected narration. Probably each operator would select a different first sentence; and that sentence, even the first phrases in it, would begin to influence the next choices. Beginnings of meaning would rapidly develop a momentum of their own, exercising stronger and stronger suggestion upon later selections. Compared with natural selection, this model, of course, has the serious flaw of employing a willful selector, whose history and purposes would necessarily influence the final product. It must equally be plain, however, that the outcome might, and I think usually would, depart widely from any design the selector first had in mind. Most of it would probably emerge as the result of a dialectical interplay between the selector's preconceptions and the suggestions evoked in him by the emerging words. I think such an experiment well worth a try, if not as science, then as experimental literature. It might mark the birth of the organic novel.

I have tried in this paper to consider all too briefly some of the conditions that have molded life here, and some of the reasons for believing that they would mold life anywhere. The nub of such an argument is to bring life within the order of nature, to see development as an orderly process, everywhere affording full play to chance, but not in any important degree accidental.

We living things are a late outgrowth of the metabolism of our Galaxy. The carbon that enters so importantly into our composition was cooked in the remote past in a dying star. From it at lower temperatures nitrogen and oxygen were formed. These, our indispensable elements, were spewed out into space in the exhalations of red giants and such stellar catastrophes as supernovae, there to be mixed with hydrogen, to form eventually the substance of the sun and planets, and ourselves. The waters of ancient seas set the pattern of ions in our blood. The ancient atmospheres molded our metabolism.

We have been told so often and on such tremendous authority as to seem to put it beyond question, that the essence of things must remain forever hidden from us; that we must stand forever outside nature, like children with their noses pressed against the glass, able to look in, but unable to enter. This concept of our origins encourages another view of the matter. We are not looking into the universe from outside. We are looking at it from inside. Its history is our history; its stuff, our stuff. From that realization we can take some assurance that what we see is real.

Judging from our experience upon this planet, such a history that begins with elementary particles, leads perhaps inevitably toward a strange and moving end: a creature that knows, a science-making animal that turns back upon the process that generated him and attempts to understand it. Without his like, the universe could be, but not be known, and that is a poor thing.

Surely this is a great part of our dignity as men, that we can know, and that through us matter can know itself; that beginning with protons and electrons, out of the womb of time and the vastnesses of space, we can begin to understand; that organized as in us, the hydrogen, the carbon, the nitrogen, the oxygen, those 16 to 20 elements, the water, the sunlight — all, having become us, can begin to understand what they are, and how they came to be.

Nature of Matter

Nature of Matter

SECOND SCIENTIFIC SESSION

The first observation of an Omega minus—a particle of strangeness minus 3—photographed in the 80-inch hydrogen bubble chamber at Brookhaven National Laboratory. [*preceding photograph*]

Introduction

MELVIN CALVIN, Chairman

IN THE FIRST scientific session we discussed the history of the universe in its many aspects, including the history of living things as we know them. Here, we will be concerned with the very nature of the matter of which those things are constructed. Our view of the nature of that matter has varied greatly, ever since man began to wonder about its nature. In this session, we will undertake an examination of the present status of our concepts of its nature, at all levels of organization.

In constructing this program, we were concerned about the order. The very structure of our experience, as Dr. Wigner will describe, in a very real sense determines in some way our view of things.

First we will concern ourselves not with any chronological sequence of our concepts either in universal time or human time, but rather with an examination of the concepts and materials at various levels of organization. We begin with the Axiom itself, which gives rise to the physical laws, and presumably there is more than one. In the very structure of these physical laws in turn we arrive at the first particles which, for lack of a better name, following consultation with some of my colleagues — both theoretical and experimental — we may call Aons, for the leptons and baryons. We proceed from these to the particles of still higher degrees of order and structure: i.e., the atoms, both the nuclei and the electronic structure surrounding them. From here we proceed to a still higher degree of order. How are the atoms hooked together into aggregations, which we call molecules? And, finally, how the molecules are put together into living structures, of which man might be considered one of the most complex representations.

Thus, we have a series of four class A contributors to this symposium, with one who discusses the Axioms, one who discusses the Aons, one who discusses the Atoms, and one who discusses the Aggregates of Atoms.

MELVIN CALVIN Laboratory of Chemical Biodynamics, University of California, Berkeley

Symmetry and Conservation Laws

EUGENE P. WIGNER

INTRODUCTION

SYMMETRY AND invariance considerations, and even conservation laws, undoubtedly played an important role in the thinking of the early physicists, such as Galileo and Newton, and probably even before them. However, these considerations were not thought to be particularly important and were articulated only rarely. Newton's equations were not formulated in any special coordinate system and thus left all directions and all points in space equivalent. They were invariant under rotations and displacements, as we now say. The same applies to his gravitational law. There was little point in emphasizing this fact, and in conjuring up the possibility of laws of nature which show a lower symmetry. As to the conservation laws, the energy law was useful and was instinctively recognized in mechanics even before Galileo.[1] The momentum and angular momentum conservation theorems in their full generality were not very useful even though in the special case of central motion they give, of course, one of Kepler's laws. Most books on mechanics, written around the turn of the century and even later, do not mention the general theorem of the conservation of angular momentum[2]. It must have been known quite generally because those dealing with the three-body-problem, where it is useful, write it down as a matter of course. However, people did not pay very much attention to it.

This situation changed radically, as far as the invariance of the equations is concerned, principally as a result of Einstein's theories. Einstein articulated the postulates about the symmetry of space, that is, the equivalence of directions and of different points of space, eloquently.[3] He also re-established, in a modified form, the equivalence of coordinate systems in motion and at rest. As far as the conservation laws are concerned, their significance became evident when, as a result of the interest in Bohr's atomic model, the angular momentum conservation theorem became all-important. Having lived in those days, I know that

EUGENE P. WIGNER Princeton University

there was universal confidence in that law as well as in the other conservation laws. There was much reason for this confidence because Hamel, as early as 1904, established the connection between the conservation laws and the fundamental symmetries of space and time.[4] Although his pioneering work was practically unknown, at least among physicists, the confidence in the conservation laws was as strong as if it had been known as a matter of course to all. This is yet another example of the greater strength of the physicist's intuition than of his knowledge.

Since the turn of the century, our attitude toward symmetries and conservavation laws has turned nearly full circle. Few articles are written nowadays on basic questions of physics which do not refer to invariance postulates and the connection between conservation laws and invariance principles has been accepted, perhaps, too generally.[5] In addition, the concept of symmetry and invariance has been extended into a new area — an area where its roots are much less close to direct experience and observation than in the classical area of space-time symmetry. It may be useful, therefore, to discuss first the relation of phenomena, laws of nature, and invariance principles to each other. This relation is not quite the same for the classical invariance principles, which will be called geometrical, and the new ones, which will be called dynamical. Finally, I would like to review, from a more elementary point of view than customary, the relation between conservation laws and invariance principles.

EVENTS, LAWS OF NATURE, INVARIANCE PRINCIPLES

The problem of the relation of these concepts is not new; it has occupied people for a long time, first almost subconsciously. It may be of interest to review it in the light of our greater experience and, we hope, more mature understanding.

From a very abstract point of view, there is a great similarity between the relation of the laws of nature to the events on one hand, and the relation of symmetry principles to the laws of nature on the other. Let me begin with the former relation, that of the laws of nature to the events.

If we knew what the position of a planet is going to be at any given time, there would remain nothing for the laws of physics to tell us about the motion of that planet. This is true also more generally: if we had a complete knowledge of all events in the world, everywhere and at all times, there would be no use for the laws of physics, or, in fact, of any other science. What I am making is the rather obvious statement that the laws of the natural sciences are useful because, without them, we would know even less about the world. If we already knew the position of the planet at all times, the mathematical relation between these positions which the planetary laws furnish would not be useful but might still be interesting. It might give us a certain pleasure and perhaps amazement to contemplate, even if it would not furnish us new information. Perhaps also, if someone came who had some different information about the positions of that planet, we could more effectively contradict him if his statements about the positions did not conform with the planetary laws — assuming that we have confidence in the laws of nature which are embodied in the planetary law.

Let us turn now to the relation of symmetry or invariance principles to the laws of nature. If we know a law of nature, such as the equations of electrodynamics, the knowledge of the subtle properties of these equations does not add anything to the content of these equations. It may be interesting to note that the correlations between events which the equations predict are the same no matter whether the events are viewed by an observer at rest, or by an observer in uniform motion. However, all the correlations between events are already given by the equations themselves and the aforementioned observation of the invariance of the equations does not augment the number or change the character of the correlations.

More generally, if we knew all the laws of nature, or the ultimate law of nature, the invariance properties of these laws would not furnish us new information. They might give us a certain pleasure and perhaps amazement to contemplate, even though they would not furnish new information. Perhaps, also, if someone came around to propose a different law of nature, we could more effectively contradict him if his law of nature did not conform with our invariance principle — assuming that we have confidence in the invariance principle.

Evidently, the preceding discussion of the relation of the laws of nature to the events, and of the symmetry or invariance principles to the laws of nature is a very sketchy one. Many, many pages could be written about both. As far as I can see, the new aspects which would be dealt with in these pages would not destroy the similarity of the two relations — that is, the similarity between the relation of the laws of nature to the events, and the relation of the invariance principles to the laws of nature. They would, rather, support it and confirm the function of the invariance principles to provide a structure or coherence to the laws of nature just as the laws of nature provide a structure and coherence to the set of events.

GEOMETRICAL AND DYNAMICAL PRINCIPLES OF INVARIANCE

What is the difference between the old and well-established geometrical principles of invariance, and the novel, dynamical ones? The geometrical principles of invariance, though they give a structure to the laws of nature, are formulated in terms of the events themselves. Thus, the time-displacement invariance, properly formulated, is: the correlations between events depend only on the time intervals between the events, not on the time at which the first event takes place. If P_1, P_2, P_3 are positions which the aforementioned planet can assume at times t_1, t_2, t_3, it could assume these positions also at times $t_1 + t$, $t_2 + t$, $t_3 + t$ where t is quite arbitrary. On the other hand, the new, dynamical principles of invariance are formulated in terms of the laws of nature. They apply to specific types of interaction, rather than to any correlation between events. Thus, we say that the electromagnetic interaction is gauge invariant, referring to a specific law of nature which regulates the generation of the electromagnetic field by charges, and the influence of the electromagnetic field on the motion of the charges.

It follows that the dynamical types of invariance are based on the existence of specific types of interactions. We all remember having read that, a long time ago, it was hoped that all interactions could be derived from mechanical interactions. Some of us still remember that, early in this century, the electromagnetic interactions were considered to be the source of all others. It was necessary, then, to explain away the gravitational interaction, and in fact this could be done quite successfully. We now recognize four or five distinct types of interactions: the gravitational, the electromagnetic, one or two types of strong (that is, nuclear) interactions, and the weak interaction responsible for beta decay, the decay of the μ meson, and some similar phenomena. Thus, we have given up, at least temporarily, the hope of one single basic interaction. Furthermore, every interaction has a dynamical invariance group, such as the gauge group for the electromagnetic interaction.

This is, however, the extent of our knowledge. Otherwise, let us not forget, the problem of interactions is still a mystery. Utiyama[6] has stimulated a fruitful line of thinking how the interaction itself may be guessed once its group is known. However, we have no way of telling the group ahead of time, we have no way of telling how many groups and hence how many interactions there are. The groups seem to be quite disjointed and there seems to be no connection between the various groups which characterize the various interactions or between these groups and the geometrical symmetry group which is a single, well-defined group with which we have been familiar for many, many years.

GEOMETRICAL PRINCIPLES OF INVARIANCE AND CONSERVATION LAWS

Since it is good to stay on terra cognita as long as possible, let us first review the geometrical principles of invariance. This was recognized by Poincaré first, and I like to call it the Poincaré group.[7] Its true meaning and importance were brought out only by Einstein, in his special theory of relativity. The group contains, first, displacements in space and time. This means that the correlations between events are the same everywhere and at all times, that the laws of nature — the compendium of the correlations — are the same no matter when and where they are established. If this were not so, it would have been impossible for the human mind to find the laws of nature.

It is good to emphasize at this point the fact that the laws of nature, that is the correlations between events, are the entities to which the symmetry laws apply, not the events themselves. Naturally, the events vary from place to place. However, if one observes the positions of a thrown rock at three different times, one will find a relation between those positions, and this relation will be the same at all points of the earth.

The second symmetry is not at all as obvious as the first one: it postulates the equivalence of all directions. This principle could be recognized only when the influence of the earth's attraction was understood to be responsible for the difference between up and down. In other words, contrary to what was just said, the events between which the laws of nature establish correlations are not the three positions of the thrown rock, but the three positions of the rock, with respect to the earth.

The last symmetry — the independence of the laws of nature from the state of motion in which it is observed as long as this is uniform — is not at all obvious to the unpreoccupied mind.[8] One of its consequences is that the laws of nature determine not the velocity but the acceleration of a body: the velocity is different in coordinate systems moving with different speeds, the acceleration is the same as long as the motion of the coordinate systems is uniform with respect to each other. Hence, the principle of the equivalence of uniformly moving coordinate systems, and their equivalence with coordinate systems at rest, could not be established before Newton's second law was understood; it was at once recognized then, by Newton himself. It fell temporarily into disrepute as a result of certain electromagnetic phenomena until Einstein re-established it in a somewhat modified form.

It has been mentioned already that the conservation laws for energy, linear and angular momentum are direct consequences of the symmetries just enumerated. This is most evident in quantum mechanical theory where they follow directly from the kinematics of the theory, without making use of any dynamical law, such as the Schrödinger equation. This will be demonstrated at once. The situation is much more complex in classical theory and, in fact, the simplest proof of the conservation laws in classical theory is based on the remark that classical theory is a limiting case of quantum theory. Hence, any equation valid in quantum theory, for any value of Planck's constant h, is valid also in the limit $h = 0$. Traces of this reasoning can be recognized also in the general considerations showing the connection between conservation laws and space-time symmetry in classical theory. The conservation laws can be derived also by elementary means, using the dynamical equation, that is, Newton's second law, and the assumption that the forces can be derived from a potential which depends only on the distances between the particles. Since the notion of a potential is not a very natural one, this is not the usual procedure. Mach, for instance, assumes that the force on any particle is a sum of forces, each due to another particle.[9] Such an assumption is implicit also in Newton's third law; otherwise the notion of counterforce would have no meaning. In addition, Mach assumes that the force depends only on the positions of the interacting pair, not on their velocities. Some such assumption is indeed necessary in classical theory.[5] Under the assumptions just mentioned, the conservation law for linear momentum follows at once from Newton's third law and, conversely, this third law is also necessary for the conservation of linear momentum. All this was recognized already by Newton. For the conservation law of angular momentum which was, in its general form, discovered almost 60 years after the *Principia* by Euler, Bernoulli, and d'Arcy, the significance of the isotropy of space is evident. If the direction of the force between a pair of particles were not directed along the line from one particle to the other, it would not be invariant under rotations about that line. Hence, under the assumptions made, only central forces are possible. Since the torque of such forces vanishes if they are oppositely equal, the angular momentum law follows. It would not follow if the forces depended on the positions of three particles or more.

In quantum mechanics, as was mentioned before, the conservation laws follow already from the basic kinematical concepts. The point is simply that the states in quantum mechanics are vectors in an abstract space and the physical quantities, such as position, momentum, etc., are operators on these vectors. It then follows, for instance, from the rotational invariance that, given any state φ, there is another state φ_α which looks just like φ in the coordinate system that is obtained by a rotation α about the Z axis. Let us denote the operator which changes φ into φ_α by Z_α. Let us further denote the state into which φ goes over in the time interval τ by $H_\tau\varphi$ (for a schematic picture, cf. Figure 1). Then, because of the

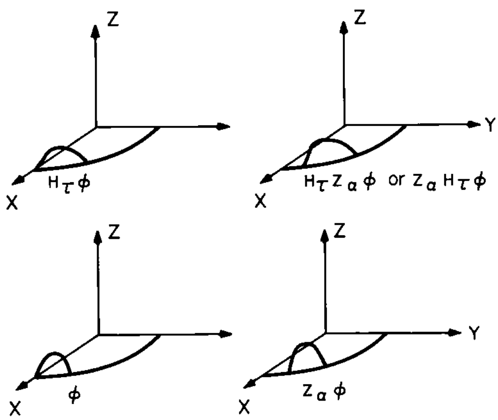


FIGURE 1

rotational invariance, φ_α will go over, in the same time interval, into the state $H_\tau\varphi_\alpha$ which looks, in the second coordinate system, just like $H_\tau\varphi$. Hence, it can be obtained from $H_\tau\varphi$ by the operation Z_α. It follows that

$$H_r Z_\alpha \varphi = Z_\alpha H_r \varphi \tag{1}$$

and since this is valid for any φ,

$$H_r Z_\alpha = Z_\alpha H_r. \tag{2}$$

Thus the operator Z_α commutes with H_r, and this is the condition for its being conserved. Actually, the angular momentum about the Z axis is the limit of $(Z_\alpha - 1)/\alpha$ for infinitely small α. The other conservation laws are derived in the same way. The point is that *the transformation operators, or at least the infinitesimal ones among them, play a double role and are themselves the conserved quantities.*

This will conclude the discussion of the geometrical principles of invariance. You will note that reflections which give rise *inter alia* to the concept of parity were not mentioned, nor did I speak about the apparently much more general geometric principle of invariance which forms the foundation of the general theory of relativity. The reason for the former omission is that I will have to consider the reflection operators anyway at the end of this discussion. The reason that I did not speak about the invariance with respect to the general coordinate transformations of the general theory of relativity is that I believe that the underlying invariance is not geometric but dynamic. Let us consider, therefore, the dynamic principles of invariance.

DYNAMIC PRINCIPLES OF INVARIANCE

When we deal with the dynamic principles of invariance, we are largely on terra incognita. Nevertheless, since some of the attempts to develop these principles are both ingenious and successful, and since the subject is at the center of interest, I would like to make a few comments. Let us begin with the case that is best understood, the electromagnetic interaction.

In order to describe the interaction of charges with the electromagnetic field, one first introduces new quantities to describe the electromagnetic field, the so-called electromagnetic potentials. From these, the components of the electromagnetic field can be easily calculated, but not conversely. Furthermore, the potentials are not uniquely determined by the field, several potentials (those differing by a gradient) give the same field. It follows that the potentials cannot be measurable and, in fact, only such quantities can be measurable which are invariant under the transformations which transform a potential into an equivalent potential. This invariance is, of course, an artificial one, similar to that which we could obtain by introducing into our equations the locations of a ghost. The equations then must be invariant with respect to changes of the coordinate of that ghost. One does not see, in fact, what good the introduction of the coordinate of the ghost does.

So it is with the replacement of the fields by the potentials, as long as one leaves everything else unchanged. One postulates, however, and this is the decisive step, that in order to maintain the same situation, one has to couple a transformation of the matter field with every transition from a set of potentials to another one which gives the same electromagnetic field. The combination of these two transformations, one on the electromagnetic potentials, the other on the matter field, is called a gauge transformation. Since it leaves the physical situation unchanged, every equation must be invariant thereunder. This is not true, for instance, of the unchanged equations of motion, and they would have, if left unchanged, the absurd property that two situations which are completely equivalent at one time would develop, in the course of time, into two distinguishable situations. Hence, the equations of motion have to be modified, and this can be done most easily by a mathematical device called the modification of the Lagrangian. The simplest modification that restores the invariance gives the accepted equations of electrodynamics which are well in accord with all experience.

Let me state next, without giving all the details, that a similar procedure is possible with respect to the gravitational interaction. Actually, this has been hinted at already by Utiyama[6]. The unnecessary complication in this case is the introduction of generalized coordinates. The equations then have to be invariant with respect to all the coordinate transformations of the general theory of relativity. This would not change the content of the theory but would only amount to the introduction of a more flexible language in which there are several equivalent descriptions of the same physical situation. Next, however, one postulates that the matter field also transform as the metric field so that one has to modify the equations in order to preserve their invariance. The simplest modification, or one of the simplest ones, leads to Einstein's equations.

The preceding interpretation of the invariance of the general theory of relativity does not interpret it as a geometrical invariance. That this should not be done, was pointed out already by the Russian physicist Fock[10]. With a slight oversimplification, one can say that a geometrical invariance postulates that two physically different situations, such as those on Figure 1, should develop in the course of time into situations which differ in the same way. This is not the case here: the postulate is merely that two different descriptions of the same situation should develop in the course of time into two descriptions which also describe the same physical situation. The similarity with the case of the electromagnetic potentials is obvious.

Unfortunately, the situation is by no means the same in the case of the other interactions. One knows very little about the weaker one of the strong interactions. The stronger one and also the weak interaction have groups which are first of all very much smaller than the gauge group or the group of general coordinate transformations.[11] Instead of the infinity of generators of the gauge and general transformation groups, they have only a finite number, that is, eight, generators. They do suffice, nevertheless, to determine the form of the interaction to a large extent, as well as to derive some theorems, similar to those of spectroscopy, which give approximate relations between reaction rates and between energies, that is, masses. Figure 2 shows the octuplet of heavy masses — its members are joined to each other by the simplest nontrivial representation of the underlying group which is equivalent to its conjugate complex.

Another difference between the invariance groups of electromagnetism and

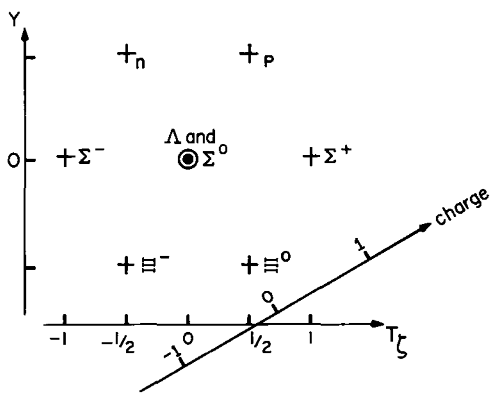


FIGURE 2

gravitation on one hand, and at least the invariance group of the strong interaction on the other hand, is that the operations of the former remain valid symmetry operations even if the existence of the other types of interactions is taken into account. The symmetry of the strong interaction, on the other hand, is "broken" by the other interactions; i.e., the operations of the group of the strong interaction are valid symmetry operations only if the other types of interactions can be disregarded. The symmetry group helps to determine the interaction operator in every case. However, whereas all interactions are invariant under the groups of the electromagnetic and gravitational interactions, only the strong interaction is invariant under the group of that interaction.

We have seen before that the operations of the geometric symmetry group entail conservation laws. The question naturally arises whether this is true also for the operations of the dynamic symmetry groups. Again, there seems to be a difference between the different dynamic invariance groups. It is common opinion that the conservation law for electric charge can be regarded as a consequence of gauge invariance; i.e., of the group of the electromagnetic interaction. On the other hand, one can only speculate about conservation laws which could be attributed to the dynamic group of general relativity. Again, it appears reasonable to assume that the conservation laws for baryons and leptons can be deduced by means of the groups of the strong and of the weak interactions.[12] If true, this would imply that the proper groups of these interactions have not yet been recognized. One can adduce two pieces of evidence for the last statement. First, so far, the conservation laws in question[13] could not be deduced from the symmetry properties of these interactions, and it is unlikely that they can be deduced from them.[14] Second, the symmetry properties in question are not rigorous but are broken by the other interactions. It is not clear how rigorous conservation laws could follow from approximate symmetries — and all evidence indicates that the baryon and lepton conservation laws are rigorous.[13] Again, we are reminded that our ideas on the dynamical principles of invariance are not nearly as firmly established as those on the geometrical ones.

Let me make a last remark on a principle which I would not hesitate to call symmetry principle and which forms a transition between the geometrical and dynamical principles. This is given by the crossing relations.[15] These relate to the probability amplitudes of any collision, such as

$$A + B + \cdots \rightarrow X + Y + \cdots \tag{3}$$

This will be a function of the invariants which can be formed from the four-vector momenta of the incident and emitted particles. It then follows from one of the reflection principles which I did not discuss, the "time reversal invariance," that the amplitude of (3) determines also the amplitude of the inverse reaction

$$X + Y + \cdots \rightarrow A + B + \cdots \tag{4}$$

in a very simple fashion. If one reverses all the velocities and also interchanges past and future (which is the definition of "time reversal"), (4) goes over into (3) so that the amplitudes for both are essentially equal. Similarly, if we denote the anti-particle of A by $\overline{A}$, that of B by $\overline{B}$, and so on, and consider the reaction

$$\overline{A} + \overline{B} + \cdots \rightarrow \overline{X} + \overline{Y} + \cdots \tag{5}$$

its amplitude is immediately given by that of (1) because (according to the interpretation of Lee and Yang), the reaction (5) is obtained from (3) by space inversion. The amplitudes for

$$\overline{X} + \overline{Y} + \cdots \rightarrow \overline{A} + \overline{B} + \cdots \tag{6}$$

can be obtained in a similar way. The relations between the amplitudes of reactions (3), (4), (5), (6) are consequences of geometrical principles of invariance.

However, one can go further. The crossing relations tell us how to calculate, for instance, the amplitude of

$$\overline{X} + B + \cdots \rightarrow \overline{A} + Y + \cdots \tag{7}$$

from the probability amplitudes of reaction (3). To be sure, the calculation, or its result, is not simple any more. One has to consider the dependence of the reaction amplitude for (3) as an analytic function of the invariants formed from the momenta of the particles in (3), and extend this analytic function to such values of the variables which have no physical significance for the reaction (3) but which give the amplitude for (7). Evidently there are several other reactions the amplitudes of which can be obtained in a similar way; they are all obtained by the analytic continuation of the amplitude for (3), or any of the other reactions. Thus, rather than exchanging A and X to obtain (7), A and Y could be exchanged, and so on.

The crossing relations share two properties of the geometrical principles of invariance: they do not refer to any particular type of interaction and most of us believe that they have unlimited validity. On the other hand, though they can be formulated in terms of events, their formulation presupposes the establishment of a law of nature, namely the mathematical, in fact analytic, expression for the collision amplitude for one of the aforementioned reactions. One may hope that they will help to establish a link between the now disjoint geometrical and dynamical principles of invariance.

Elementary Particles

GEOFFREY F. CHEW

IN MY SURVEY I am going to borrow freely from an article currently being prepared in collaboration with Murray Gell-Mann and Arthur Rosenfeld. In the title of our article we have used the term "particle" but have carefully avoided the adjective "elementary." Had the title of this paper been of my choosing, I should similarly have avoided such a modifier, because if there is a discernible trend in current high energy physics it is the *decline* of the elementary particle concept. A major purpose of my report to you is to describe this decline of an ancient idea and the beginnings of a new idea that may fill the void.

PARTICLE PROLIFERATION

One well-publicized facet of the present position in subatomic physics is the proliferation of nuclear particles. The number of particles discovered during the past 5 years is so large that there is little point in an enumeration here. Those of you not already familiar with the list would only be bewildered, not enlightened. What was the source of the spectacular increase in the rate of particle discovery?

GEOFFREY F. CHEW Department of Physics and Lawrence Radiation Laboratory, University of California, Berkeley

There were at least three ingredients. To understand these you must realize that most nuclear particles are unstable and have rest mass energies of the order of a billion electron volts. Some with longer lifetimes were found in cosmic rays but the majority are so short-lived that they must be created in the laboratory if they are to be seen. Accelerators in the billion-volt range became operative during the last 10 years; this was a necessary first ingredient for the population explosion among established particles.

A second ingredient was the invention and development of the bubble chamber. This marvelous instrument allowed the complete study of almost all the particles produced in a given nuclear reaction, under conditions where thousands of events could be accumulated in a reasonable time. Unstable particles with lifetimes as short as 10^{-22} sec. thereby revealed themselves through their decay products, which could be followed in detail and analyzed on a statistical basis. The third ingredient was the development of high-speed data analysis systems capable of handling tens of thousands of events in individual experiments. Without such systems the output from bubble chambers would overwhelm human capacities.

By no means all of our information about nuclear particles is coming from bubble chambers. Spark chambers and sophisticated electronic arrangements continue to be essential in situations in which the number of events to be analyzed runs in the hundreds of thousands or more. A substantial number of the new particles have revealed themselves in such experiments.

What can be said to summarize the vast amount of particle information now available? First of all, proliferation has occurred only among "strongly interacting particles." These are particles like the neutron and proton that mutually interact through powerful short-range forces. All particles discovered to date, with the exception of the photon, electron, muon, and the neutrinos are of this type. The latter four particles are called leptons and this lepton family has not increased in number over the past 30 years. Recently it was discovered at Brookhaven that there are two kinds of neutrinos, one associated with muons and one with electrons, but this cannot be considered proliferation in the sense observed for strongly interacting particles.

A second generalization is that there are no strongly interacting particles of mass very small compared to the average, which is $\backsim$ 1 billion electron volts in order of magnitude. The lightest is the pion whose mass is $\backsim$ 1/7 Bev. Compare this to the electron whose mass is $\backsim 5 \times 10^{-4}$ Bev or to the neutrino whose mass is believed to be zero. A plausible inference from these two circumstances is that strongly interacting particles are in some sense dynamical structures, owing their existence to the same forces through which they mutually interact. Such a mechanism can be imagined to produce a spectrum of energy levels (i.e., particle masses) of the type observed, with no sharp upper bound, and with the ground state not differing qualitatively from excited states.

In contrast, the origin of the leptons cannot be dynamical in the same sense. The necessary forces here have never been found and the bizarre spectrum of states, beginning with one of zero energy and terminating sharply at the muon, bears no resemblance to any dynamical spectrum seen before. There have been many discoveries of simplicity and regularity in lepton properties in recent years but nothing yet to suggest a deep explanation of why these particles exist. I shall concentrate the remainder of this report, therefore, on the strongly interacting particles — where the accumulated weight of experimental evidence promises soon to produce a major increment of understanding.

It will be easy for you to grasp one reason for the current optimism regarding strong interactions. The point is that the very magnitude of the particle population is transforming high energy physics into something like spectroscopy, as it was at the beginning of the century. Regularities are being discovered that allow the particles — or energy levels — to be grouped into families. Some of these groupings already have been given a dynamical motivation by the theorists, others remain at present empirical. But if the current rate of experimental discovery continues, it seems inevitable that major aspects of the underlying fundamental principles will reveal themselves.

YUKAWA FORCES

A second reason for optimism is that hundreds of painstaking measurements over the past 30 years have by now convincingly verified the Yukawa hypothesis concerning the forces acting between strongly interacting particles. Expressed in modern language this hypothesis associates nuclear forces with the existence of anti-particles through the key concept of "crossing." Consider a reaction leading from the two-particle "channel" (a,b) to the channel (c,d):

$$a + b \leftrightarrow c + d,$$

the probability of the reaction occurring being given by the absolute square of the reaction amplitude $A(E_a, E_b, E_c, E_d)$ which is a function of the four particle-energies. The principle of crossing states that this same function also describes two "crossed" reactions that correspond to replacing ingoing particles by outgoing anti-particles and vice versa:

$$a + \bar{c} \leftrightarrow \bar{b} + d,$$
$$a + \bar{d} \leftrightarrow \bar{b} + c.$$

These three different reactions are distinguished by the signs of the energy variables, which are positive or negative according to whether ingoing or outgoing particles are involved, but if the controlling amplitude is known for one reaction it can be obtained for the two others by smooth extrapolation (continuation) in energy.

An example of crossing is the following pair of reactions involving neutrons and protons

$$(a) \qquad n + p \leftrightarrow n + p$$
$$(b) \qquad n + \bar{n} \leftrightarrow p + \bar{p},$$

both described by the same reaction amplitude, an important part of which is represented by Figure 1. The first way of drawing the arrows in this diagram

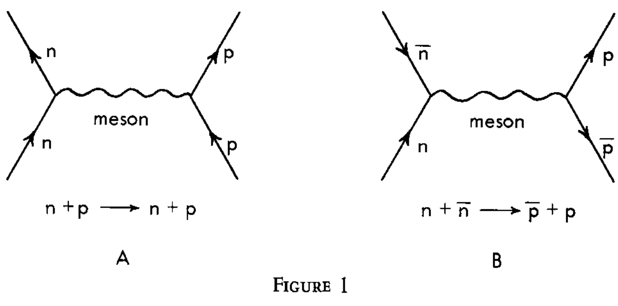


FIGURE 1

(which physicists will recognize is related to a Feynman diagram, though it has not quite the same meaning) is appropriate to reaction (a) and the second to reaction (b). The two figures differ, of course, only in the direction we read them, as indicated by the arrowheads.

The second way of drawing the figure can be interpreted as the contribution from the meson "intermediate state" in reaction (b):

$$n + \bar{n} \rightarrow \text{meson} \rightarrow p + \bar{p}.$$

In general, whenever there exists a particle whose quantum numbers permit it to "communicate" with both sides of a reaction the contribution from such a diagram as (*b*) is necessarily present in the reaction amplitude. This contribution is given by the so-called Breit-Wigner formula and depends on the mass (or energy) of the intermediate state and the coupling between this state and the initial and final channels. Physicists refer to such coupling strengths as "partial widths."

The first way of drawing Figure 1 has an entirely different interpretation. Here we say that a meson is "exchanged" in a scattering collision between a neutron and a proton and it can be shown that this exchange constitutes a "force" acting between these particles. This, of course, is the Yukawa force, its range and magnitude being determined by the Breit-Wigner parameters. In general we may say that forces arise from the exchange of particles that can act as intermediate states for crossed reactions. If the masses and partial widths of these particles are known, the details of the forces can be predicted.

I have already stated that there has been accumulated ample experimental verification for the existence of these Yukawa forces, with the predicted strength and range. It is important to add, however, that no evidence for *additional* forces has been found. It is possible to believe, in other words, that *all* the forces between strongly interacting particles arise by the Yukawa mechanism. This situation should be contrasted with that of conventional electrodynamics where the photon interacts by direct absorption and emission in a manner that cannot be described as due to the exchange of particles communicating with crossed reactions. (It is, of course, true that electromagnetic interactions between *charged* particles *are* of the Yukawa type, arising from the exchange of photons.) In the strongly interacting family we see no indication of particles subject, like the photon, to direct emission and absorption. For many years it was believed that the π-meson might be such a particle, but experiments indicate nothing unusual about the forces felt by the π. Suggestions have been made that the spin-one mesons may have a special status, but again there is nothing in the data to support such an idea.

NUCLEAR PARTICLE DEMOCRACY

I am led now to a third reason for optimism about strong interactions. This is the success achieved in understanding the existence and properties of certain particles, assuming them to be dynamical composites (bound states) of other particles, held together by the Yukawa forces. A well-known example is the deuteron, which in a first approximation can be considered a neutron-proton composite held together by the exchange of various mesons. It is of great importance to realize, however, that the deuteron is not exactly composed of one proton plus one neutron. In quantum mechanics it is more accurate to say that *most of the time* the deuteron consists of these two particles. The deuteron communicates with a great variety of channels in addition to $p + n$, and according to quantum theory any state consists part of the time of each of the channels that communicate with it. If one neglects other channels such as $n + p + \pi^{\circ}$ in a calculation of deuteron properties, one does not get an exact result. Nevertheless there is a general belief that, since the simplest channel successfully accounts for the bulk of the observed deuteron properties, it should eventually be possible to improve the predictions systematically by inclusion of more channels.

A similar situation holds with respect to the Δ particle (sometimes called the 3, 3 resonance), whose properties have been qualitatively understood assuming it to be a composite of pion plus nucleon, held together primarily by exchange of a nucleon. The predictions here are cruder than for the deuteron because the neglected channels play a larger role. For most particles, in fact, a single-channel approximation is not adequate even to achieve qualitatively correct results. Many channels must simultaneously be included, which poses a grave calculational task. The problem of including all the significant channels is in most cases still too difficult for us, but suppose we could solve it. The question is would we then get a correct description of each particle? Would the quantum numbers and the mass come out right? Or do we need to put into our calculation some *special extra* parameters pertaining to certain particles? Until recently there was an almost universal belief that a few strongly interacting particles, such as the nucleon, would have to play a special role; but during the past 2 years the notion of *democracy* among strongly interacting particles has been gaining adherents, that is to say, the notion that no strongly interacting particle is more elementary than any other.

The revolutionary character of nuclear particle democracy is best appreciated by contrasting the aristocratic structure of atomic physics as governed by quantum electrodynamics. No attempt is made there to explain the existence and properties of the electron and the photon; one has always accepted their masses, spins, etc., together with the fine structure constant, as given parameters. There exist atomic particles, such as positronium, whose properties are calculable in the sense described above, but so far one does not see a plausible basis, even in principle, for computing the properties of photon and electron as we compute those of positronium. In particular the zero photon mass and the small magnitude of the fine structure constant appear most unlikely to emerge from dynamics of the Yukawa type.

What has happened to make some of us think that all strongly interacting particles are composite, with properties that are dynamically calculable? I have already mentioned the systematic absence of very small masses. For a long time we have known of certain particles, such as the deuteron and Δ, for which there has been qualitative success in calculating the properties from Yukawa forces; but a presumed analogy with electrodynamics inhibited theorists from attempting similar calculations for the nucleon — which from the time of its discovery had been accorded a status parallel to that of the electron. Gradually, however, it was realized that this select status was dubious — that no observed properties of the nucleon justify the belief that it differs in a fundamental way from other strongly interacting particles. And so, finally, an attempt was made to calculate nucleon properties from Yukawa forces: the same qualitative success was achieved as for Δ! The status of these two particles, N and Δ, now appears completely parallel. It seems, furthermore, on the basis of recent theoretical developments, that in all such dynamical calculations no distinction need be made on the basis of the spin or other quantum numbers of the particle involved. To sum up, if there is no need for aristocracy among strongly interacting particles, may there not be democracy?

Once suggested, the notion of nuclear particle democracy appears plausible, but one cannot yet say that it is established. Some physicists continue to believe, for example, that certain spin-one mesons are not ordinary citizens but play a role like that of the photon. The most promising experimental tests of the democracy principle involve careful measurement of nuclear reactions at very high energies, but currently available energies from the Brookhaven and CERN accelerators are insufficient to make such a test decisive. A further increase by at least a factor of five is required, necessitating the construction of still larger machines.

Even if accepted, the democracy principle leaves unanswered some major questions. One is why the photon and the leptons are excluded. Perhaps one should take comfort here in the historical circumstance, emphasized recently by Dirac, that up to now nature has revealed her secrets in a remarkably well-separated sequence of installments and that human intellect has been able to grasp only one installment at a time. Why nature should be so considerate to scientists no one understands, but she has been so in the past and at the moment she seems to be inviting us to understand strong interactions as a more or less isolated collection of phenomena. A unified picture of the entire subatomic world may not appear until a later and quite separate installment.

Many physicists feel unhappy about the prospect of a separate theory for strong interactions because Yukawa's idea about nuclear forces arose in the first place from a presumed analogy with electromagnetism, as also did Fermi's picture of the weak interaction. By the same token, however, there are many points of analogy between electromagnetism and *gravitation*, and nearly all of us believe it is too soon to invoke general relativity in subatomic physics. Eventually such a a step will be taken, but human science proceeds one step at a time. How many separate steps will be required it is impossible to predict.

BOOTSTRAP DYNAMICS

I conclude this report by mentioning a second major question, this one of immediate concern: the origin of the special symmetries and the associated conservation laws that characterize strong interactions. I am thinking here of isotopic spin and strangeness as well as the newly discovered eightfold way. For those of you hearing these terms for the first time, it is sufficient to say that they describe an empirical classification of the observed particles into families, all the members of a given family having similar properties. The analogy with the atomic periodic table is evident. No explanation has yet been given for the symmetries underlying these subatomic families but many physicists believe that the secret will emerge from the "bootstrap" mechanism.

The bootstrap concept is equivalent to the notion already developed of a democracy governed by Yukawa forces. Each strongly interacting particle is conjectured to be a bound state of those channels with which it communicates, owing its existence entirely to forces associated with the exchange of particles that communicate with "crossed" channels. Each of these latter particles in turn owes *its* existence to a set of forces to which the original particle makes a contribution. In other words, each particle helps to generate other particles, which in turn generate it. In this circular and violently non-linear situation it is possible to imagine that no free parameters appear and that the only self-consistent set of particles is the one we find in nature. Needless to say, vigorous efforts are being made to investigate this possibility.

Now if the system is in fact self-determining, perhaps the special strong interaction symmetries are not arbitrarily to be imposed but will emerge as necessary components of self-consistency. Perhaps their origin is destined to be understood at the same moment we understand the pattern of masses and spins for strongly interacting particles — *both* aspects of the system flowing from the dynamics of the bootstrap.

On this optimistic note I close my survey.

The Structure of Nuclei

VICTOR F. WEISSKOPF

THE PREVIOUS lecture presented a vivid picture of the subnuclear world of new phenomena and particles, a world which manifests itself when matter is exposed to energies of more than 100 MeV per atom. The topics of the following discussions deal with phenomena which involve energy exchanges of less than 100 MeV. We are getting less ambitious and closer to our human environment here on earth. Under these conditions, the list of elementary particles can be reduced to three: the proton and the neutron, on the one hand, and the electron on the other. Perhaps the neutrino should be included here too, but it plays only a very ephemeral role.

It is one of the blessings of quantum mechanics that we can completely forget about effects of the internal structure of a unit when dealing with phenomena, involving energies much lower than the excitation threshold of this unit. We can

VICTOR F. WEISSKOPF Commission Européenne pour la Recherche Nucléaire, Geneva, Switzerland

forget about atomic structure in kinetic gas theory, about nuclear structure in most of atomic theory, and about mesons and hyperons in most of nuclear theory.

In the study of nuclear structure, we are interested in bound systems — the atomic nuclei — formed by neutrons and protons under the influence of a predominantly attractive nuclear force. The nuclei, in turn, form much larger and weaker bound systems with electrons — the atoms — under the influence of the electric force. There is no discussion on atomic structure scheduled in these series; it is assumed that this structure is well known. It might be appropriate therefore to compare nuclear structure with atomic structure by stressing the similarities found in these two fundamental units of matter. In both cases we deal with systems whose properties are based upon the binding effects of fields upon particles. As Newton wrote 260 years ago, "There are therefore agents in nature able to make the particles of bodies stick together by a very strong attraction. And it is the business of experimental philosophy to find them out."

We follow Newton's appeal and turn to those agents. Since we can neglect gravity here, we face two forces between the three elementary particles under consideration — the electromagnetic and the nuclear force. The first figure contains a comparison of some salient features of these two fields of force. The sources of the electric field are the electrons and protons; the sources of the nuclear field are the neutrons and protons. The strength of these sources is determined by a quantity "charge" whose square is most conveniently measured in units of the product hc of Planck's quantum of action and the light velocity. Thus, the electric source strength is given by the value $e^2/hc = 1/137$. The corresponding nuclear source

FIELD

	electric	nuclear
Coupling constant	$\dfrac{e^2}{hc} = \dfrac{1}{137}$	$\left(\dfrac{g^2}{hc}\right)_{\mathrm{eff}} \sim 0.1$
Potential (main part)	$\dfrac{e^2}{r}$ attractive repulsive	$\dfrac{g^2}{r} e^{-r/r_0}$ $\quad r_0 \sim 1.4 \times 10^{-13}$ cm mainly attractive

FIGURE 1

strength cannot be defined in such simple terms. The structure of the nuclear field seems to be much more complicated, in particular when high energy phenomena and small distances are involved. It is possible, however, to express the effects of the nuclear forces between protons and neutrons in terms of a potential, when one restricts the description to effects with relatively low energy transfer. Even this potential is a complicated function of the distance and the relative spin and parity; the most important term, however, which governs the nuclear force potential at larger distances, has a form not too different from the electrostatic Coulomb-potential e^2/r. It is the so-called Yukawa potential $(g^2/r)e^{-r/r_0}$. This allows us to introduce a quantity g, which can be considered as a "nuclear charge" and which determines the strength of the nuclear field in analogy to the electric charge e. The corresponding source strength g^2/hc amounts to about 1/10.

The dominant part of the potential for the two binding forces is therefore of similar structure. The electric force can be attractive or repulsive, depending on the sign of the charges, the nuclear force is mainly attractive except for a strong repulsion at very small distances (less than 0.4×10^{-13} cm). The nuclear Yukawa potential is characterized by a range r_0 of 1.4×10^{-13} cm. The force decreases exponentially beyond this range, whereas the electric potential reaches to infinity. The small range of the nuclear force causes an important difference. One can build up a macroscopic electric field by assembling a large surplus of charge in a given volume since the field reaches to large distances. Nucleons, however, cannot be compressed into mutual distances of less than the range r_0. Hence, a macroscopic nuclear field can never be realized.

Attractive forces between particles give rise to bound systems. Let us compare atomic and nuclear systems of few particles (see Figure 2). The simplest are, of course, the two-particle systems: the hydrogen atom in the electric case and the deuteron in the nuclear realm. The size and binding energy of these systems are determined by the laws of quantum mechanics; the ground state will be the state of lowest possible energy. There is an interplay of two factors which determines the situation: On the one hand, the potential energy is lower at smaller distances; on the other hand, the kinetic energy increases in quantum systems when the dimension of the system becomes smaller. The minimum is reached for the H-atom at dimensions of the order of a Bohr-radius a and the corresponding binding energy is given by the Rydberg Ry. Both quantities are made up of

FEW - PARTICLE - SYSTEMS

	electric	nuclear
System:	H-atom, light atoms	Deuteron, light nuclei
Size:	Bohr-radius $a = \frac{\hbar^2}{me^2} = 0.5 \times 10^{-8}$ cm	Nuclear Bohr-radius $a_N = \frac{\hbar^2}{Mg^2} \approx 2 \times 10^{-13}$ cm
Binding energy:	Rydberg $Ry = \frac{me^4}{2h^2} = 13$ ev	Nuclear Rydberg $Ry^N = \frac{Mg^4}{2h^2} \approx 5$ MeV

FIGURE 2

fundamental magnitudes such as charge and mass of the electron and the quantum of action.

The situation in simple nuclear systems should not be much different. The nuclear potential energy has a similar dependence on the distance between particles as the electric one. Hence, size and binding energy of nuclear systems should be determined by corresponding quantities which we will call the nuclear Bohr-radius a_N and the nuclear Rydberg Ry_N. They are the same quantities as the atomic ones, except that the electric charge e is replaced by the nuclear charge g, and the electron mass m by the nucleon mass M. There is, of course, a difference between the two dominant potentials and that is the finite range of the nuclear force. However, this range is of the same order as a_N so that the finite range effects do not alter the situation in any essential way. Within the range, the nuclear force is not qualitatively different from the Coulomb force. The deuteron and other light nuclei have in fact the dimension of a nuclear Bohr-radius, and the binding energies are of the order of a nuclear Rydberg. Thus the ratio of nuclear to atomic dimensions is given by $(m/M)(e^2/g^2) \backsim 3 \times 10^{-5}$. The ratio of the relevant energies is

$$(M/m)(g^4/e^4) \backsim 4 \times 10^5.$$

There is one characteristic difference which is introduced by the finite range of nuclear forces: the H-atom possesses excited states of many different quantum numbers; the deuteron does not have any excited states. The nuclear forces are not reaching far enough to bind two particles in states of higher quantum numbers.

The effects of the finite range become more pronounced when we go on to consider systems with a large number of particles (see Figure 3). On the electric side we should consider two types, complex atoms and molecules; on the nuclear side we look at complex nuclei. The decisive factor in complex atoms is the strong charge of the nucleus. Its overriding attraction reaches all electrons and compresses them in a small volume. This compression is proportional to the central charge:

MANY - PARTICLE - SYSTEMS

	electric		nuclear
System:	Complex atoms	Molecules	Complex nuclei
Density	$\frac{Z}{a^3}$ electrons	$\frac{1}{a^3}$ atoms independent of number	$\frac{1}{a_N^3}$ nucleons independent of number
Binding energy:	$Z^{4/3} \cdot Ry$	$\sim Ry$ (less)	$\sim Ry_N$
Structure:	Orbits in common potential (I.P.M.)	Rigid frame of nuclei	Orbits in common potential (I.P.M.) $Z \sim N$ $A < 250$

FIGURE 3

The electron density in complex atoms therefore is roughly proportional to the atomic number.

In a complex nucleus, however, the nuclear force of a single nucleon is not strong enough for holding a large number of particles within the range of its force. There will be interactions only between a few neighboring particles. Consequently, we obtain an assembly with a density independent of the total number of particles. The average binding energy per particle reflects the same conditions: in the atom it increases strongly with the number of particles; in the nucleus it is roughly independent of it and is of the order of a nuclear Rydberg.

The dynamical structure of the complex atom and the complex nucleus shows remarkable similarities, however. In both cases, the particles arrange themselves so as to minimize the total energy; this leads to a spread of the wave functions over the largest possible area, within the space available at equilibrium density.

Hence, the particles appear to be "smeared out" and do not assume definite locations. This leads to a very characteristic situation which allows the application of the so-called Independent Particle Model (I.P.M.) in which each particle moves almost as if it were a single particle embedded in a potential well produced by the average position of the others. We know the consequences of this in the atomic and nuclear realm: the dynamic situation can be described approximately as a superposition of single-particle orbits.

It is interesting to consider the molecular structure in this connection. The chemical forces between atoms are also short-ranged and repulsive at small distances and therefore produce systems in which the density of units and the binding energy are roughly independent of the number; however, two types of particles of very different mass are involved — nuclei and electrons; this fact brings about a most peculiar effect. Only the light particles spread over the volume available, the heavy ones are subject to the same potential energy, and therefore their spread will be much smaller because of the smaller zero point kinetic energy of a heavier particle when subject to the same potential. Hence, they assume well-defined locations and establish that framework which is so characteristic for the molecular structure and the structure of solid state.

Let us return to the dynamic properties of nuclei and atoms. They can be described by a combination of simple particle orbits in a common potential; a central Coulomb-like potential in one case, a potential well in the other (see Figure 4). This description immediately leads one to expect a shell structure, as it

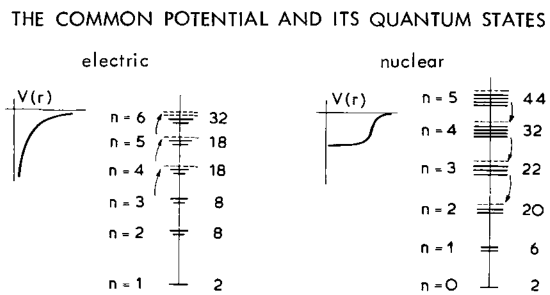


FIGURE 4. Spectrum of single particles in average potential of atoms and nuclei. Shell structure is seen. The arrows indicate the levels that are shifted from one shell to the other because of shielding and spin orbit effects.

is observed both in atoms and in nuclei. But the number of particles in the different shells depends on the type of potential and on other detailed features. In general, each shell corresponds to a given value of the main quantum number; in the atomic case the shielding effects move certain states of higher angular momentum into the next higher shell, whereas in the nuclear case a strong spin orbit coupling does essentially the opposite.

It is interesting to compare the properties of systems with not completely filled shells. Here the repulsion between electrons and the attraction between nucleons produce different effects. The electrostatic repulsion favors orbits that overlap as little as possible. Hence, the occupied orbits tend to stay apart and form a combination which displays approximate spherical distribution. In the nuclei, however, the nucleons occupy orbits which overlap as much as possible, and this often produces deformed nuclei, since a large number of particles would favor orbits of similar asymmetry. This collective deformation of unfilled shells is a characteristic phenomenon in nuclear structure.

The common shell structure in both systems leads to many similarities in the spectra of atoms and nuclei in their systematics and in their structure. The deformed nuclei, however, exhibit spectral properties which have their analogue rather in the spectra of molecules of similar symmetry, in particular in respect to rotational and vibrational motions.

In view of this discussion it is not surprising that nuclear spectroscopy presents many analogies to the atomic one. Transitions between different levels take place with the emission of light quanta, such as in atoms. The emission of "nuclear field quanta" — the mesons — is not possible since the energy differences are of the order of a nuclear Rydberg, which is much less than the rest mass of the lightest meson. There is, however, another form of transition possible in nuclei: it is the emission of a lepton pair, an electron or positron, and a neutrino. They are known as radioactive beta-decay of nuclei and are very slow processes, much slower than the emission of light. Such transitions involve, of course, a change of charge of the nucleus. But nuclear forces are charge-independent, and therefore the charge plays a very small role in the structure of nuclei. Hence there are many analogies between γ-ray emission and lepton-pair-emission, apart from the different over-all rate in time. One part of the interaction of nuclei with lepton pairs — the vector-interaction part — is strictly proportional to the interaction with electromagnetic fields, according to recent most interesting investigations.

Another way to observe transitions between nuclear states is the analysis of collision processes, in which nuclei are bombarded by nuclear particles. These processes — nuclear reactions — are analogous to atomic excitation or ionization by electron impacts. The corresponding nuclear processes are in fact better and more thoroughly investigated. They represent the most important source of information about nuclear structure. One classes the effects of nuclear collisions in two groups: the first one contains those in which the projectile interacts only with one or very few nuclear particles; they are called "direct reactions." The other group contains collision processes in which the energy of the incoming particle is distributed over a large part of the nuclear constituents. These are "compound-reactions" and they lead to secondary effects, which can be described successfully by a heating process of the compound system, followed by an evaporation of one or more particles.

In many respects, nuclear structure and nuclear reaction are better known than the corresponding atomic properties. This might seem surprising in view of the much smaller size of nuclear systems and in view of our incomplete knowledge regarding nuclear forces. It comes to a large extent from the fact that nuclei are well shielded by the electron shells of the atom from any disturbing influences, whereas atoms are usually investigated under conditions in which they are strongly disturbed by the environment. A nucleus can spend hours in an excited state, if the angular momentum law excludes its de-excitation. An atom in a similar state would be de-excited quickly in the laboratory by collision with neighbors. Hence, the nuclear symmetries and the ensuing conservation laws are more readily established. Our knowledge of structure and dynamics is always based upon recognition of the relevant symmetries and conservation laws.

Let us summarize the significant points which emerge in the study of nuclear structure:

1. A new field of force makes its appearance, which has no macroscopic realization, but which nevertheless is of fundamental importance in the structure of matter. It produces an attraction between nucleons, which is similar in character and in its effects to the electric attraction between nuclei and electrons.
2. Nuclear physics represents a second field of validity for non-relativistic quantum physics. The world within the nucleus presents us with a surprising repetition of phenomena encountered in the atomic world; a

repetition with strikingly similar features such as shell structure, orbitals, and the familiar systematics of spectra. Nature repeats itself here on a higher scale in energy, and on a smaller scale in dimensions. An impressive illustration of that correspondence is the energy production by binding of smaller units into larger ones: the parallelity between chemical combustion yielding about 1 ev per unit, and nuclear combustion yielding about 1 MeV per unit.

3. Nuclear physics is of special importance for certain natural phenomena in the universe. Most of the phenomena on earth are determined by the laws of atomic physics, since the temperatures on the earth's surface are too small for nuclear energies. The nuclear phenomena which we study in the laboratory are mostly man-made. Hence the natural place of nuclear processes is in the center of stars, where they are essential for energy production and element formation. It is symbolic for the ever-increasing scope of science that in our nuclear research we have recreated here on earth a world of phenomena whose natural place is outside of our abode.

The Architecture of Molecules

LINUS PAULING

MOLECULAR ARCHITECTURE may be said to have originated in 1874, when J. H. van't Hoff and J. A. le Bel independently formulated the brilliant postulate that the four valence bonds of the carbon atom are directed approximately toward the corners of a regular tetrahedron. It was extended into inorganic chemistry in 1893, when A. Werner suggested that in many inorganic complexes six atoms are arranged at the corners of a regular octahedron about a central atom, and that other geometrical structures are represented by other complexes. However, the present-day subject of molecular architecture, involving the precise discussion of

LINUS PAULING Gates and Crellin Laboratories of Chemistry, California Institute of Technology

the structure of molecules and crystals in terms of interatomic distances and bond angles, is a product of the last half-century. It began in 1913, when W. H. Bragg and W. L. Bragg reported the determination of the arrangement of atoms of sodium and chlorine in the sodium chloride crystal and the evaluation of the distance between the centers of the atoms by the newly discovered method of analysis of the x-ray diffraction pattern. During the last 50 years the precise structures of thousands of crystals and molecules have been determined by the methods of x-ray diffraction, electron diffraction, and molecular spectroscopy, with the aid of other techniques, such as the determination of entropy values.

The structures of many crystals and molecules can be conveniently described by reference to the five regular polyhedra — the tetrahedron, the octahedron, the cube, the icosahedron, and the pentagonal dodecahedron. Thus the crystal beryllium oxide has a structure involving tetrahedral coordination: each beryllium atom has as its nearest neighbors four oxygen atoms, which surround it at the corners of a regular tetrahedron. In the sodium chloride crystal each atom of one kind is surrounded by six atoms of the other kind, which lie at the corners of a regular octahedron, and in the cesium chloride crystal each atom of one kind is surrounded by eight atoms of the other kind, at the corners of a cube.

The classical structure theory of organic chemistry was developed in the years following 1874 on the basis of the postulate of the tetrahedral carbon atom, which accounted for the existence of enantiomeric pairs of organic substances, with the power of rotating the plane of polarization of light by equal amounts in opposite directions. The tetrahedral character of the carbon atom was then verified in a striking way in 1914 by W. H. Bragg and W. L. Bragg, when they determined the structure of diamond and found that each carbon atom in the diamond crystal is surrounded by four other carbon atoms at the corners of a regular tetrahedron. I continue to be astounded that during the 40-year period between 1874, when the postulate of the tetrahedral carbon atom was proposed, and 1914, when the Braggs reported their determination of the structure of diamond, no scientist had described this tetrahedral arrangement of atoms and suggested that it might represent the structure of diamond. An English investigator, W. Barlow,[1] had in 1883 described several ways that seemed reasonable to him for packing atoms together in crystals, and had correctly assigned one of them to sodium chloride, another to cesium chloride, and others to certain cubic and hexagonal metals. Barlow and his collaborator Pope assigned diamond also to a close-packed arrangement, with ligancy 12, rather than 4, for the carbon atom. Their failure to have discovered the diamond arrangement may have resulted from their lack of confidence in the postulate of the tetrahedral carbon atom.

In the molecule of methane, CH_4, the four hydrogen atoms lie close to the corners of a regular tetrahedron, and the bond angles HCH have the value 109° 28′ characteristic of a regular tetrahedron. In order to account qualitatively for the optical activity of a molecule in which the four atoms attached to the carbon atom are different, such as fluorochlorobromomethane, HCFClBr, it is not necessary to assume that the bonds are directed toward the corners of a regular tetrahedron, but only that the four atoms attached to the carbon atom not lie in a plane with the carbon atom. Many experimental studies of the structure of molecules of this sort have been carried out since 1930, and it has in fact been found that the tetrahedral carbon atom is nearly regular, with the bond angles equal to the tetrahedral angle 109° 28′ to within better than 3°, in most molecules.

During the period from 1916 to 1930 it was customary for chemists, following the lead of Gilbert Newton Lewis, to describe atoms other than carbon, such as the oxygen atom in the water molecule, as tetrahedral: the oxygen atom was described as having four electron pairs in its outer shell, two unshared pairs and two pairs involved in forming the bonds to the hydrogen atoms, and the four pairs were assumed to be located at tetrahedron corners. In 1931 J. C. Slater[2] pointed out that the quantum mechanical theory of the electronic structure of molecules suggests that the bond angle should be approximately 90°. He assigned one of the unshared pairs of electrons to the 2*s* orbital of the oxygen atom, which has a lower energy value than the 2*p* orbital. The bonding electrons of the oxygen atom are then described as occupying bond orbitals that are largely 2*p* in character, and Slater pointed out that these 2*p* orbitals have their maximum values in directions in space at 90° to one another, and that the bond angle should accordingly be approximately 90°. In the water molecule the HOH bond angle is 104.5°, and in H_2S, H_2Se, and H_2Te the bond angles lie between 92° and 90°.

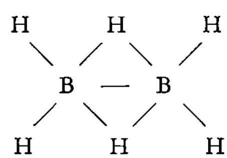
The theory of the tetrahedral carbon atom leads to the description of two carbon atoms that form a double bond with one another as two tetrahedra sharing a common edge. If the tetrahedra were regular the value of the angle between a single bond and the double bond would be 125° 16′. Many experimental values for this angle lie in the range 124 to 126°; for example, the value of the $C-C=C$ angle in propylene, $H_3C-CH=CH_2$, is 124° 45′ ± 20′. Ethylene,

$H_2C = CH_2$, is an exception; the observed value for the angle $H - C = C$ in this molecule is 121° 20′. The corresponding angle for an atom with an unshared electron pair, such as the nitrogen atom in nitrosyl chloride, $Cl - \ddot{N} = \ddot{O}$:, is smaller in value, usually about 113°.

A useful refinement of chemical structure theory was made in 1937 by Kemp and Pitzer,[3] who discovered that the rotation of the two ends of a molecule that are connected to one another by a carbon-carbon single bond is not free, but is restricted, with the potential hump (for ethane) approximately 3 kcal/mole. The stable configuration of the ethane molecule and related molecules is the one in which the bonds are staggered, rather than eclipsed. This structural feature has been found to have great value in the interpretation and prediction of thermodynamic properties of substances, as well as their spatial configurations.

An interesting development during the last few years has been the recognition of the importance of the icosahedron and pentagonal dodecahedron in molecular architecture. The icosahedron has been found to play an important part in the structure of many electron-deficient substances, including metals and alloys.

In the discussion of the electronic structure of many molecules and crystals it is found that each of the stable orbitals in each atom is occupied by an unshared electron pair or by the electrons involved in a bond. For example, in the methane molecule each hydrogen atom uses its sole stable orbital, the 1*s* orbital, in forming a bond with the carbon atom, and the carbon atom uses its four stable outer orbitals, the four tetrahedral orbitals formed by hydridization of the 2*s* orbital and the three 2*p* orbitals, in the formation of the four bonds to the hydrogen atoms. In the water molecule the four orbitals of the oxygen atom are used for formation of two bonds and for occupancy by two unshared electron pairs. In each of these molecules the number of electrons available is great enough to permit utilization of all the stable orbitals in this way. In borane, BH_3, on the other hand, there are only three electron pairs (aside from the pair of 1*s* electrons of the boron atom), so that the four stable orbitals of the boron atom cannot all be occupied. In fact, borane exists as the dimer, B_2H_6, rather than as the monomer, BH_3. The structure of B_2H_6, diborane, as determined experimentally, can be represented by the diagram



The boron atom has as its near neighbors four hydrogen atoms and the other boron atom: its ligancy is 5. Two of the hydrogen atoms, the bridging hydrogen atoms, have ligancy 2, rather than the normal value 1. The diborane molecule is called an electron-deficient molecule, because the number of valence electrons is less than the number of stable orbitals that might be used in bond formation. In diborane, as in other electron-deficient substances, the ligancy of the atoms is not only greater than the number of valence electrons, but even greater than the number of stable orbitals.

The structure of one of the more complex boranes, decaborane, $B_{10}H_{14}$, was determined by Kasper, Lucht, and Harker[4] in 1950. This structure, shown in Figure 1, is closely related to the icosahedron (see Figure 5*a*). The ten boron atoms occupy ten of the twelve corners of a nearly regular icosahedron. Ten of the hydrogen atoms lie at the corners of a larger icosahedron, each being bonded to one boron atom. The other four hydrogen atoms are bridging hydrogen atoms, each with ligancy 2. Each of the boron atoms has ligancy 6. The positions of the bridging hydrogen atoms may be described by saying that each of the two corners of the icosahedron that are not occupied by boron atoms is split between two hydrogen atoms. The observed bond lengths in the icosahedral framework have approximately the values expected for half-bonds; that is, for bonds involving electron occupancy half as great as for a shared-electron-pair bond.

Structure determinations have been made for several other complex boranes, and most of them have been found to be related to the icosahedron. Much of the recent work in this field has been carried out by W. N. Lipscomb and his collaborators. An especially interesting structure is that of the $B_{12}H_{12}^{--}$ ion, in the crystal $K_2B_{12}H_{12}$; this ion is found to have the boron atoms at the corners of a nearly regular icosahedron, with a hydrogen atom attached to each boron atom.[5]

Metals and intermetallic compounds are electron-deficient substances, with the number of valence electrons per atom less than the number of stable orbitals. As with other electron-deficient substances, the ligancy of the metal atom is greater than the number of valence electrons or of stable orbitals. The molecular architecture of many metals and intermetallic compounds has been determined by x-ray diffraction, and the structures have for the most part been found to be such as to correspond to the maximum ligancy permitted by spatial considerations; that is, the structures can be described as closest packing of spheres of the same size or, in the case of unlike atoms, of different sizes.

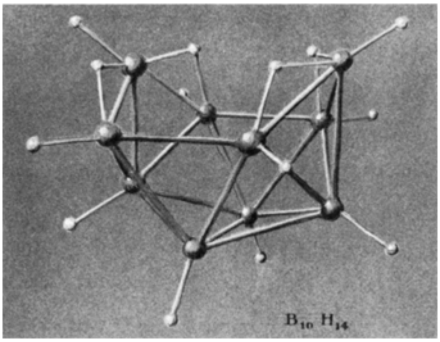


FIGURE 1. The structure of decaborane. The ten boron atoms, represented by the larger spheres, lie at ten of the twelve corners of an approximately regular icosahedron. There are four bridging hydrogen atoms and ten non-bridging hydrogen atoms.

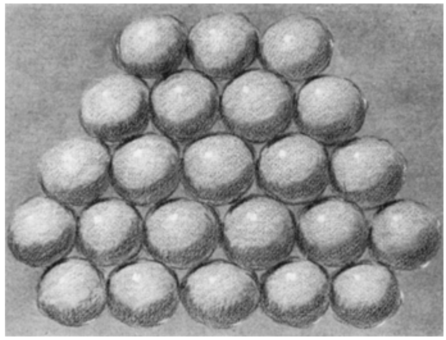


FIGURE 2. A portion of a close-packed layer of spherical atoms.

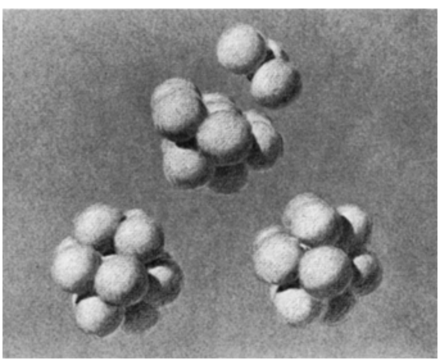


FIGURE 3. The coordination of twelve spherical atoms about a central one, as found in cubic closest-packing (lower left) and hexagonal closest-packing (lower right).

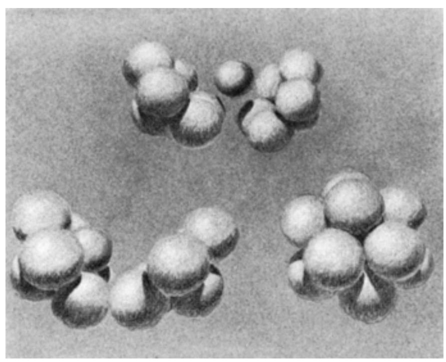
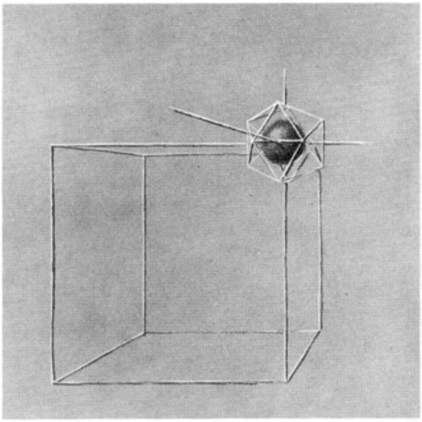
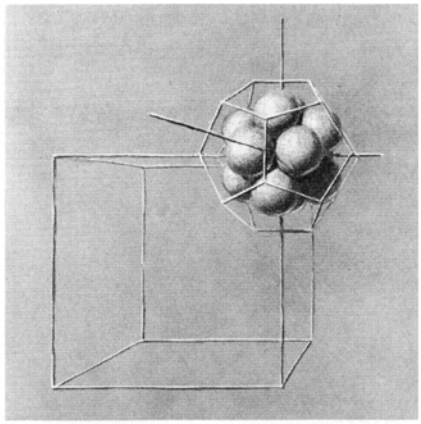
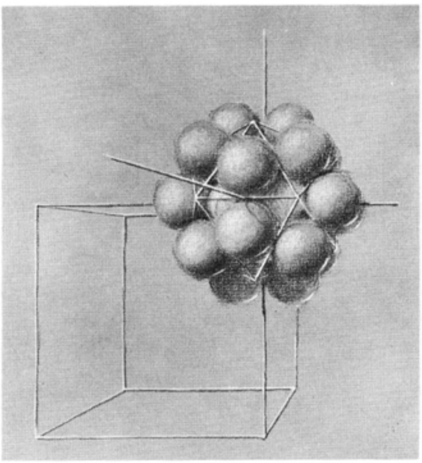


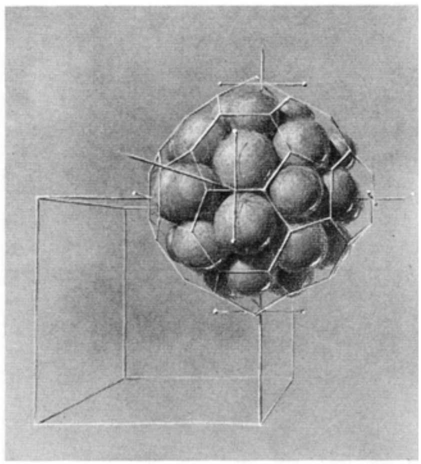
FIGURE 4. The coordination of twelve spherical atoms about a central atom with 10 per cent smaller diameter. The twelve larger atoms lie at the corners of a regular icosahedron. This is the closest possible packing of twelve equivalent atoms about a central atom.



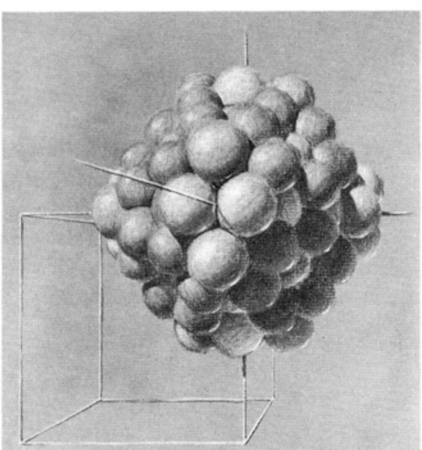


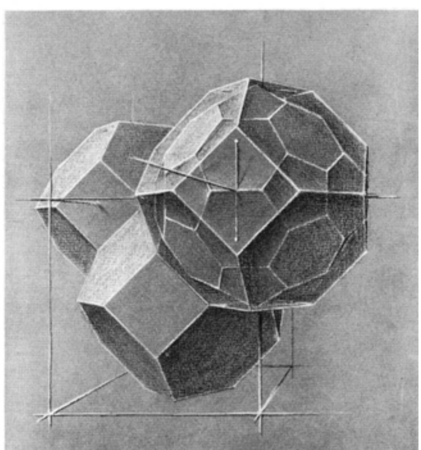
a　　*b*





c　　*d*





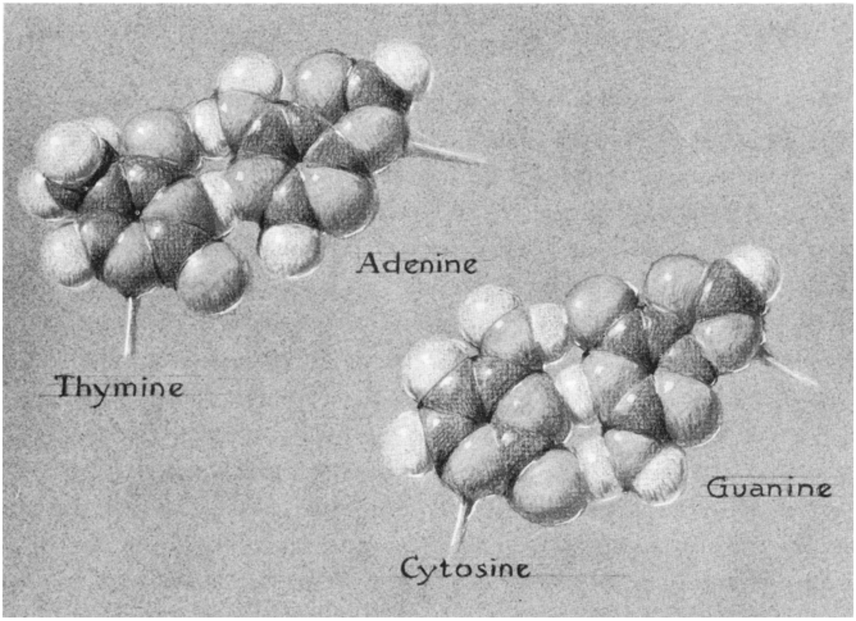


FIGURE 6. The structure of the thymine-adenine pair, forming two hydrogen bonds, and the cytosine-guanine pair, forming three hydrogen bonds, as postulated in the Watson-Crick theory of DNA.

FIGURE 5*a*. This figure and Figures 5*b* to 5*f* represent the body-centered cubic structure of 162 atoms in the unit cube of the intermetallic compound $Mg_{32}(Al, Zn)_{49}$. In the drawing an atom at a lattice point is represented as at the center of a regular icosahedron, which is outlined.

FIGURE 5*b*. The corners of the outlined pentagonal dodecahedron lie out from the centers of the 20 triangular faces of the icosahedral shell of 12 atoms.

FIGURE 5*c*. An icosahedron is shown with its corners lying out from the centers of the pentagonal faces of the dodecahedron.

FIGURE 5*d*. The completed shell of 32 atoms is indicated. The truncated icosahedron has its 60 corners out from the centers of the triangular half-faces of the rhombic triacontahedron, and 12 other atomic positions are shown, up from the centers of the hexagonal faces of the truncated icosahedron.

FIGURE 5*e*. The complex of atoms about a lattice point has 72 atoms in its outer shell, and 45 in the inner shells. The atoms of the outer shell lie on the faces of a cubo-octahedron.

FIGURE 5*f*. The complexes shown in Figure 5*e* share the atoms of their outer shells, when they are arranged about the points of a body-centered cubic lattice, as indicated here.

Most metals crystallize with one or the other or both of the two simple closest-packed structures described by Barlow, cubic closest packing and hexagonal closest packing. Each of these structures is composed of hexagonal layers of atoms, as shown in Figure 2. These layers can be superimposed in such a way that each atom in one layer is in contact with three atoms of the next layer. This superposition can be carried out in an infinite number of ways, including two in each of which the atoms are all equivalent. In all these structures the ligancy of each atom is 12. The arrangement of the twelve atoms around the central atom to which they are ligated is shown in Figure 3 — for cubic closest packing at the lower left and for hexagonal closest packing at the lower right. The coordination polyhedron in each case has six triangular faces and four square faces.

Although cubic closest packing and hexagonal closest packing (and the more complex ways of superimposing the hexagonal layers) represent the ways of packing spheres of equal size to obtain the maximum density, the complexes shown in Figure 3 do not provide the most closely-packed arrangement of 12 equivalent spherical atoms around a central atom. The smallest complex of 12 equivalent spherical atoms around a central atom is shown in Figure 4. In this complex the central atom is 10 per cent smaller than the 12 surrounding atoms, and the 12 atoms lie at the corners of a regular icosahedron.

It has been found by experiment that many intermetallic compounds have structures involving the centered icosahedron. Most of these structures are very complicated. It might be thought that the high symmetry of the icosahedron would permit it to be present in crystals with simple structure; but the symmetry elements of the icosahedron (15 twofold axes, 10 threefold axes, and 6 fivefold axes, as well as planes of symmetry) are not especially helpful in crystallization, because crystals cannot have fivefold axes of symmetry.

The intermetallic compound Mg_{32} $(Zn, Al)_{49}$ may be described as an example of an electron-deficient substance in which each atom forms the largest number of bonds with adjacent atoms permitted by the relative atomic sizes. The atoms of magnesium are somewhat larger than those of zinc and aluminum, permitting icosahedral coordination. The structure of the crystal seems to be based upon the principle that each coordination polyhedron is a triangular polyhedron — all its faces are triangles. In Figure 5*a* a single (small) atom is indicated at the corner of the unit cube, with an icosahedron outlined about it. The twelve larger atoms occupying the corners of the icosahedron are shown in Figure 5*b*. The pentagonal dodecahedron that is sketched about the complex has its corners at the positions out from the centers of the triangular faces of the icosahedron. The twenty atoms occupying these positions are shown in Figure 5*c*, and a larger icosahedron is sketched in. When the twelve atoms of this larger icosahedron are introduced, the outer shell of the complex involves 32 atoms, lying at the corners of a rhombic triacontahedron. Continued application of the structural principle leads to the shell of 72 atoms shown in Figures 5*d* and 5*e* — 60 atoms at the corners of a truncated icosahedron, and 12 additional atoms out from the centers of 12 of the 20 hexagonal faces of the truncated icosahedron. These large complexes, with their centers at the corners and the centers of the unit cube, are then condensed together in the way shown in Figure 5*f*. Each of the 72 outer atoms in the outermost shell of each complex is shared between two complexes, so that each outer shell contributes 36 atoms per lattice point, which with the inner shells of 45 atoms gives 81 atoms per lattice point, a total of 162 in the unit cube.

Some even more complex structures have been determined recently. The most striking one is that of the cubic crystal with approximate formula $NaCd_2$, for which the structure determination was made by S. Samson.[6] This crystal has 1192 atoms in the cubic unit of structure. Each atom is closely surrounded by 12, 13, 14, 15, or 16 ligands; 528 of the 1192 atoms are at the centers of icosahedra.

Among the most interesting problems of science are those of the structure and properties of substances of biological importance. In this branch of molecular architecture, which has been developing rapidly during recent years, a structural feature called the hydrogen bond is of great importance.

The hydrogen bond was discovered by Latimer and Rodebush[7] in 1920. It is a bond by a hydrogen atom between two electronegative atoms. Sometimes, as in the hydrogen difluoride ion, $(FHF)^-$, the hydrogen atom lies midway between the two electronegative atoms, but usually it is attached more strongly to one than to the other. For example, in a crystal of ordinary ice each water molecule is attached to four other water molecules, which surround it in an approximately tetrahedral arrangement, by four hydrogen bonds, with two of the hydrogen atoms at 1.00 A and the other two (which are strongly bonded to the other oxygen atoms) at 1.76 A, the oxygen-oxygen distance being 2.76 A. Determinations of the crystal structure of amino acids and simple peptides carried out during a period of years beginning in 1938, especially by R. B. Corey[8] and his collaborators, showed that the molecules of these substances, which are closely related to proteins, attach themselves to one another by forming hydrogen bonds between the nitrogen atom and the oxygen atom of peptide groups, with N-H . . . O distance about 2.90 A, and also showed that the peptide group is essentially planar. This knowledge led to the formulation of the alpha helix, a configuration of polypeptide chains that has been found to be present in many proteins. The great developments in the field of protein structure that have been carried out during recent years by J. B. Kendrew and M. Perutz, who have used x-ray diffraction to make nearly complete structure determinations of crystals of the globular proteins myoglobin and hemoglobin, cannot be described adequately in a brief review of molecular architecture.

The importance of the hydrogen bond in determining the structures and properties of the nucleic acids was recognized by J. D. Watson and F. H. Crick when they made their great contribution to science in formulating their helical structure for deoxyribonucleic acid.[9] This structure involves a detailed complementariness of two intertwined polynucleotide chains. The complementariness results from the formation of hydrogen bonds between a pyrimidine residue in one chain and a purine residue in the other chain, for each pair of nucleotides in the chains. The pairs are thymine-adenine and cytosine-guanine, with the first pair forming two hydrogen bonds and the second pair three,[10] as shown in Figure 6. The simplicity of the explanation of the self-duplicating property of DNA that is provided by the Watson-Crick structure gives a basis for hope that additional studies in the field of the molecular architecture of living organisms will lead ultimately to a thorough and satisfying understanding of all the phenomena of life.

The Organization of Living Matter

GEORGE E. PALADE

INTRODUCTION

LIVING SYSTEMS consist of relatively few common chemical elements,[1] yet have unique properties which set them aside from the rest of matter: they are able to control their exchanges with the surrounding medium and to respond to changes therein; they can transform energy and metabolize matter; and especially they are capable of self-duplication. These remarkable properties, expressed in one form or another, have intrigued and challenged man's mind at least since the beginning of history. The answers have varied over the ages, but for a long time have retained an element of reverence and have assumed the involvement of extraordinary forces. Now, after millennia of illusions, doubts, and probing, it turns out that these unique properties are due to the way in which common chemical elements are put together in time and space. For our times, life — human life included — is the outcome of an elaborate organization based on trivial ingredients and ordinary forces. Historically speaking, this has been a drastic readjustment which is still affecting, sometimes with devastating force, whole fields of human endeavor.

THE STRUCTURAL HIERARCHY OF LIVING SYSTEMS

The elaborate organization just mentioned consists of a hierarchy of structural patterns which begins with relatively simple molecules and proceeds, step by step — first to polymeric macromolecules; then to molecular or macromolecular aggregates, which often take the form of elementary structures[2] such as fibrils, membranes, and particles; then to aggregates of such structures which turn out to be cell organs; then to cells; and beyond cells to tissues, organs, and organisms.

Life with its full complement of attributes, life self-sustained and self-reproducible emerges at the cell level within the hierarchy. The sufficiency of this

GEORGE E. PALADE The Rockefeller Institute

stage is attested by the fact that so many living forms do not go beyond it, and by the ease with which cells of metazoic origin revert to an independent state under culture *in vitro*. Below this level there are self-reproducible forms, like the viruses, but they are not self-sustained; they subsist only as cell symbionts or cell parasites. Hence our discussion can be limited to living matter at the cellular and subcellular level.

Already at this level, life depends on an extensive organization in depth, on a superimposition of patterns which amount to infinitely more order than matter usually tolerates. This thermodynamically improbable situation is achieved by continuously supplying energy to a pre-existing structural framework. It is this framework, the hierarchy of patterns already mentioned, that captures energy and matter from outside sources and channels them into a complex series of reactions whose outcome is the maintenance of the system and its eventual duplication. As far as we know, at present the framework is always inherited, never created *de novo*. How it originally came into being is a fascinating problem which has tempted many minds, without as yet obtaining a satisfactory answer.

THE EXTENT OF CELL STRUCTURE AND ITS RELATION TO METABOLISM

Before analyzing the structural framework of the cell, its extent should be defined and its position in relation to cell metabolism should be understood. It is usually assumed that cell structure is limited to those elements sturdy enough to resist biochemical or morphological preparative procedures. As a rule these elements consist of matter in a solid or mesomorphic state. Undoubtedly, extensive disorder is introduced by the procedures mentioned wherever cell organization relies on weaker and fewer forces. Even matter in relatively rapid transit through a cell is disposed at any given time with a certain amount of order imposed by the rates of the reactions in which it is involved. Finally water, quantitatively the major component of living systems, is by necessity extensively organized in relation to various cell structures. Hence we are dealing with a gradient of order which encompasses practically the whole cell, and of which only parts survive our attempts to analyze the whole.

As far as relations to metabolism are concerned, it should be understood that structure and metabolism are two aspects of the same set of phenomena. Matter flows continuously through a living system, it is metabolized by it and, while in transit, forms the hierarchy of patterns on which the system depends. What the biologist describes as structure are usually the more slowly revolving phases of the general metabolism. Atoms, molecules, and even larger aggregates move continuously in and out of the framework, but since their replacement is gradual, the patterns subsist while their components turn over at rates that vary throughout the hierarchy. The only exception appears to be the genome of the cell or, in other words, its set of DNA[3] molecules, which may last untouched throughout a succession of cell generations.

THE STRUCTURAL FRAMEWORK OF THE CELL

General considerations: If we consider the structural framework of the cell, as we know it today, we are faced with a rather disturbing situation: the information about it is already staggering in volume, yet far from complete and, therefore, far from being properly understood. During the last decades, however, enough progress has been made in the analysis of two different structural levels, namely those of macromolecules and of subcellular structures, to justify a tentative interpretation of the whole framework.

At the macromolecular level, the primary, secondary, and tertiary structure of many proteins, the structure of DNA and various types of RNA's, and that of other biological polymers has been elucidated to a greater or lesser extent, with remarkable and far-reaching results. One of them is the fundamental importance assumed by certain structural features of these macromolecules, primarily by their surface details, viewed in terms of tridimensional configuration as well as distribution of functional groups. Such details apparently constitute the genetic code in the DNA molecule; the template for protein synthesis at the surface of messenger RNA; the adapter that recognizes the right spot on the template for the corresponding amino acid in the case of transfer RNA; the active sites at the surface of protein molecules endowed with enzymatic properties, and so on. This level in the hierarchy of patterns turns out to be not only fundamental, but also common, as far as we know at present, for all living systems. All apparently rely on the same type of molecules and the same kind of structural details to transmit genetic information from one generation to another, and to give expression to this information in the synthesis and activity of their specific constituents.

The other level at which substantial progress has been recorded is that of cellular and subcellular structure. Here, however, the situation is different: the corresponding patterns are no longer common to all living systems, and the functional significance of many of their features remains unknown. Yet the cellular level deserves full attention for, as already mentioned, the fundamental characteristics of life emerge only at this level, not before it.

Recent advances in the analysis of cellular organization have been achieved mainly by two relatively new technical developments: electron microscopy and cell fractionation procedures. The first technique means resolving power of the order of 10–5 Å, an improvement of $\backsim$200–400 times over light microscopy, and the possibility of carrying structural analysis down to the molecular level of organization. The second allows the isolation in mass of most subcellular structures and their subsequent chemical and functional characterization *in vitro*. By an unusually fortunate coincidence, the two developments have been contemporary; hence, they have been frequently combined, and have led to an integrated, structural and functional characterization of many cell organs.

The results obtained can be put into better perspective by considering first the average dimensions and main features of the whole cellular framework. The usual animal or plant cell is a very small body: its average diameter is $\simeq$20 μ, and its volume $\simeq$5,000 μ^3. It is bounded by a thin membrane ($\simeq$8 mμ) and its content, or protoplasm, consists of a nucleus $\simeq$5 μ in diameter ($\simeq$65 μ^3 in volume) surrounded by a shell of cytoplasm. It is in this exiguous space that the most significant part of the structural hierarchy of living matter must be fitted. This extreme miniaturization is probably imposed by the slow means of communication, primarily diffusion, that operate in the protoplasm which, as a first approximation, might be considered as an aqueous sol-gel system. At closer scrutiny, however, it turns out to be highly heterogeneous and extensively compartmented. Its cytoplasmic layer contains, for instance, numerous bodies which belong to different classes among which one of the best known is that of mitochondria: spherical or rodlike bodies with an average volume of 0.8 μ^3 and a population density of a few hundred per cell. This class of subcellular components can serve as a first example of the type of results obtained and progress made by the concurrent use of electron microscopy and cell fractionation procedures.

Mitochondria: Figure 1, taken from an article published in 1900 by Michaelis,[4] shows mitochondria stained in living cells with Janus green, a dye which when oxidized has a blue-green color. Michaelis' drawing shows clearly mitochondria as they appear in the light microscope in the cells of a gland, namely, the pancreas. This figure served as "official portrait" of the mitochondria until 12 years ago. Figure 2 shows an electron micrograph of a thin section through a part of a single mitochondrion in a similar cell. The salient point is the existence of a large amount of membranous material which forms more or less regularly arranged infoldings,[5] called cristae. Figure 3 shows a fragment of a single crista prepared by a different procedure which gives a negative image by embedding the specimen in an opaque matrix. It demonstrates the existence of small particles ~80–100 A in diameter at the surface of the cristae. These are the mitochondrial "elementary particles" recently discovered by Fernández-Morán.[6]

Shortly before their fine structure was unraveled, mitochondria were isolated in mass[7] and found to contain the enzymatic equipment the cell uses to oxidize various intermediates of the Krebs cycle,[8, 9] pyruvate and fatty acids.[9, 10] This equipment includes the terminal electron transport chain from dehydrogenases to oxygen[11] and the enzymes and factors which use the energy released during oxidation to synthesize ATP,[12] the general fuel of living systems. At that time, a correlation between concentration of oxidative enzymes and frequency of cristae was already discussed.[5, 13] During the last years, after the discovery of the "elementary particles," attempts were made to give this correlation a more precise form: recent evidence[14] suggests that the electron transport chain and its associated phosphorylating enzymes are located in the elementary particles. It is still debated whether such a particle can accommodate the entire multienzyme system involved, or only part of it[15] and in this case what particular part. But these questions are now under active investigation in many laboratories and answers will be soon forthcoming.

The main function of the mitochondria is now well understood: they supply most of the energy (~90 per cent) required for the various endergonic reactions carried out by the cell, in the form of ATP produced during oxidative phosphorylation. In addition, they are apparently involved in the energy dependent accumulation of various ions and as such in the regulation (or homeostasis) of the internal medium of the cell.[16]

The typical organization of a mitochondrion comprises an outer membrane, an inner membrane whose infoldings form the cristae, and two chambers. It is the inner membrane and its cristae which are lined with elementary particles.[17]

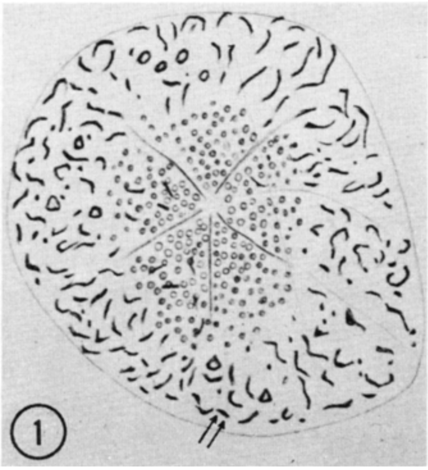
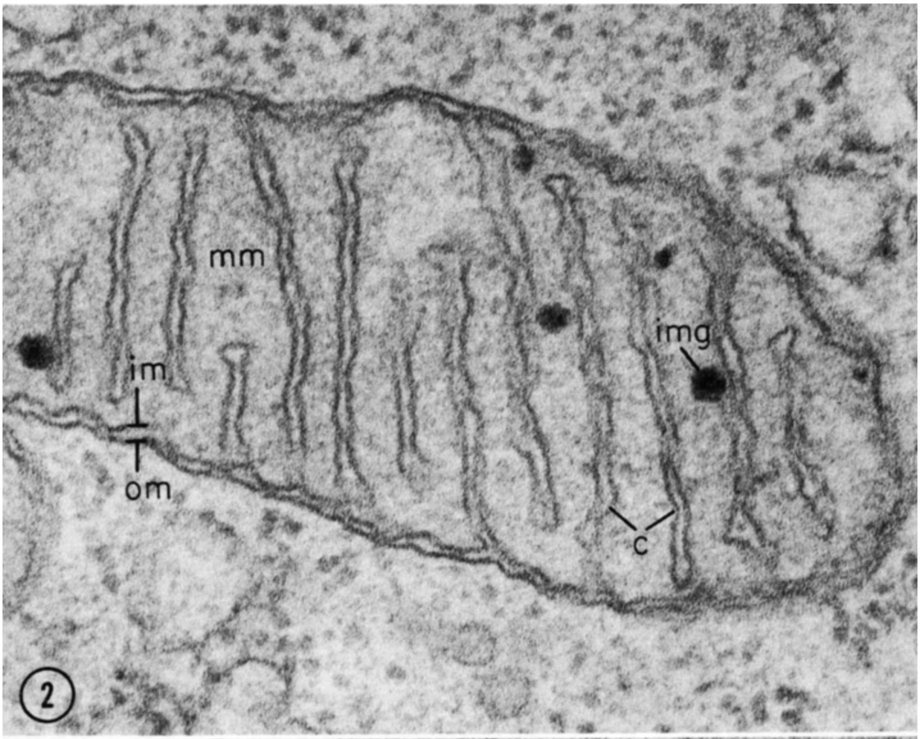


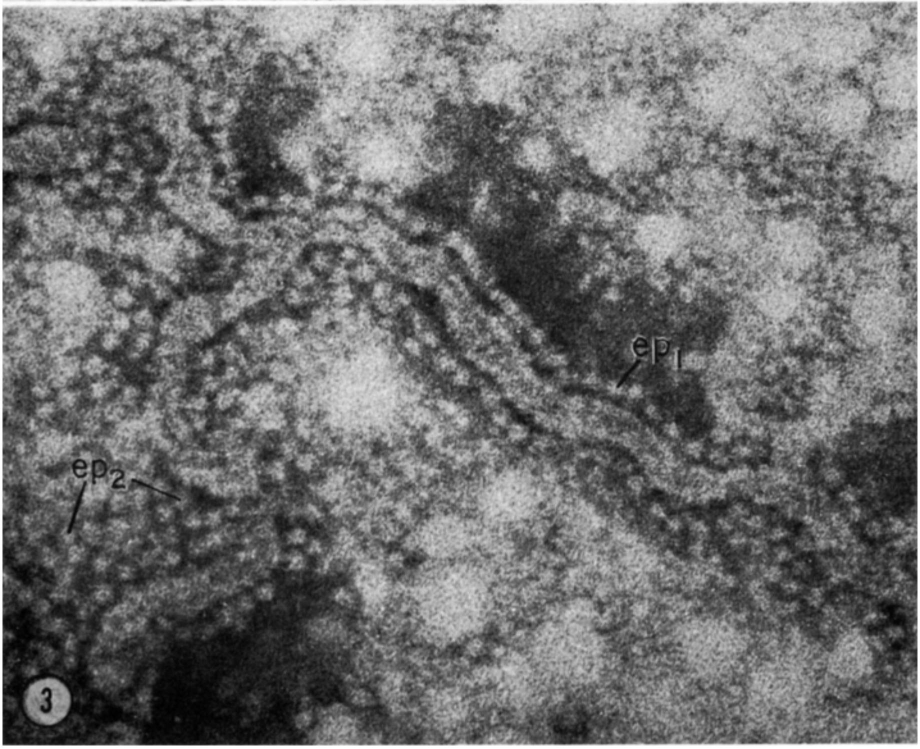
FIGURE 1. Reproduction in black and white of Figure 6 from Michaelis' paper in *Archiv mikroskop. Anat. Entwicklungsmech.*, **55**, 558 (1900). Pancreatic acinus of a mouse. The arrows (added by me) indicate part of a mitochondrion comparable to the one shown in Figure 2.

NOTE: Figures 2, 5–8, and 10–13 are electron micrographs of cells fixed in 1 per cent OsO_4, either in acetate veronal or phosphate buffer (pH 7.4–7.6), and embedded in epon. The sections were stained with lead hydroxide or uranyl acetate followed by lead hydroxide. The specimen in Figure 3 (mitochondrial fraction isolated from *Neurospora crassa*) was stained negatively with K phosphotungstate. The specimens for Figures 14 and 15 were prepared by Kellenberger's procedure. Abbreviations for all figures are as follows: *cm,* cell membrane; *n,* nucleus; *cy,* cytoplasm; *m,* mitochondrion; *er,* endoplasmic reticulum; *rs,* rough-surfaced element of the endoplasmic reticulum; *ss,* smooth-surfaced element of the endoplasmic reticulum; *ar,* attached ribosome(s); *fr,* free ribosome(s).

FIGURE 2. Electron micrograph of a section through part of a mitochondrion (pancreatic exocrine cell, guinea pig). *om:* outer membrane; *im:* inner membrane; *c:* crista; *mm:* mitochondrial matrix (inner chamber); *img:* intramitochondrial granule. × 100,000.

FIGURE 3. Electron micrograph of a negatively stained mitochondrial crista. Elementary particles lining a branching crista appear in side view at ep_1 and in full-face view at ep_2. × 300,000.


mm
im
om
img
c
2


ep1
ep2
3

Some of the features of the mitochondrial pattern can be understood in terms of available data: the outer membrane probably controls mitochondrial permeability,[18] whereas the inner membrane provides support for the multiplicity of enzymes involved in oxidative phosphorylation. Ion accumulation occurs in the inner chamber,[19] possibly in relation to some intramitochondrial granules.[20] The functional meaning of the other features remains for the moment obscure.

Although recent and still under development, the work on mitochondria has acquired historical significance for, in addition to the elucidation of the function of these subcellular components, it has introduced a number of new concepts among the basic premises with which we are trying to understand cell organization. The first concerns the existence of functionally specialized cell organs; the second recognizes the ample use of membranous material to separate compartments within the cell and to provide a solid framework for the precise, stabile positioning of active parts: each mitochondrion is a membrane-bounded cytoplasmic compartment, which in turn is subcompartmented by an inner membrane of its own. Finally, the third concept concerns the widespread occurrence of common patterns of organization at the subcellular level: mitochondria were found to have basically the same structure from animal cells to fungi.

Endoplasmic reticulum and ribosomes: The next examples will show how these concepts have influenced further research and how in turn they have been modulated by subsequent experience.

Figure 4 illustrates the distribution of "chromidia," "ergastoplasm,"[21] or "basophil substance" in a glandular cell (the exocrine cell of the pancreas), as seen by an American cytologist, A. Mathews,[22] around 1900. Electron micrographs of certain regions (i.e., basal, centrospheric, and apical [see diagram accompanying Figure 4]) of the same type of cell, which differ from one another in the extent of their "basophilia," i.e., their RNA content, are shown in Figures 5–8. The intensely basophilic basal region is packed with membrane-bounded tubules, vesicles, and cisternae (Figure 5) which belong to the intracellular membranous system of the endoplasmic reticulum.[23] Most of these elements are rough-surfaced, i.e., bear attached dense particles — the extensively studied ribosomes — on the outer surface of their limiting membranes. In addition, the intervening cytoplasm contains a sizable population of free ribosomes.[24]

The centrosphere region contains the Golgi complex,[25] a characteristic accumulation of vesicles, vacuoles, and stacked cisternae, which are smooth-surfaced, i.e., free of attached ribosomes (Figure 6). At the periphery of this complex, elements partly covered and partly free of ribosomes join the rough- to the smooth-surfaced channels of the system (Figure 7). Finally in the apical region (Figure 8) large secretion granules, bounded by a smooth-surfaced membrane, occur interspersed with ribosome-bearing elements of the endoplasmic reticulum.

It is known at present that these membrane-bounded spaces are extensively interconnected within the cell; that they probably stretch as a network from the nucleus, which is surrounded by a perinuclear cisterna,[26] to the cell membrane; and that they open at the cell surface, but that these openings as well as the communications from one part of the reticulum to another are intermittent. The cell apparently controls, at the level of the Golgi complex and of the cell membrane, the time, the volume, and the direction of flow of the content of the system.[27]

As far as we know at present, the content consists either of matter ingested in bulk from the surroundings, or substances produced by the cell to be discharged into the external medium, or both. Intermittent connections enable the cell to handle in small packets (or quanta) its products of secretion; to concentrate them for storage; and to quantitate their discharge. In the case of ingested matter, temporary isolation in a vacuole or compartment appears to be a prerequisite for concentration and digestion. The cell discharges into such a vacuole hydrolytic enzymes apparently stored in small granules or vesicles called lysosomes,[28] and thus the vacuole becomes a site of intracellular digestion. The import of matter in mass is ubiquitous among animal cells and can be considered a consequence of their heterotrophy: the fact that they depend on exogenous organic compounds to compensate their biosynthetic deficiencies. The development of the endoplasmic reticulum seems to be connected, at least in part, with such activities, but once acquired the system has been adapted to other parallel or alternative functions such as secretion.

One is tempted to assume that originally the smooth-surfaced part of the endoplasmic reticulum was primarily connected with the import of matter in bulk, while the rough-surfaced part was concerned with the production of hydrolytic enzymes, the connection between the two parts being established by a lysosome, i.e., a small storage vacuole filled with hydrolytic enzymes. In a crude analogy the smooth part of the system could be regarded as the functional equivalent of a

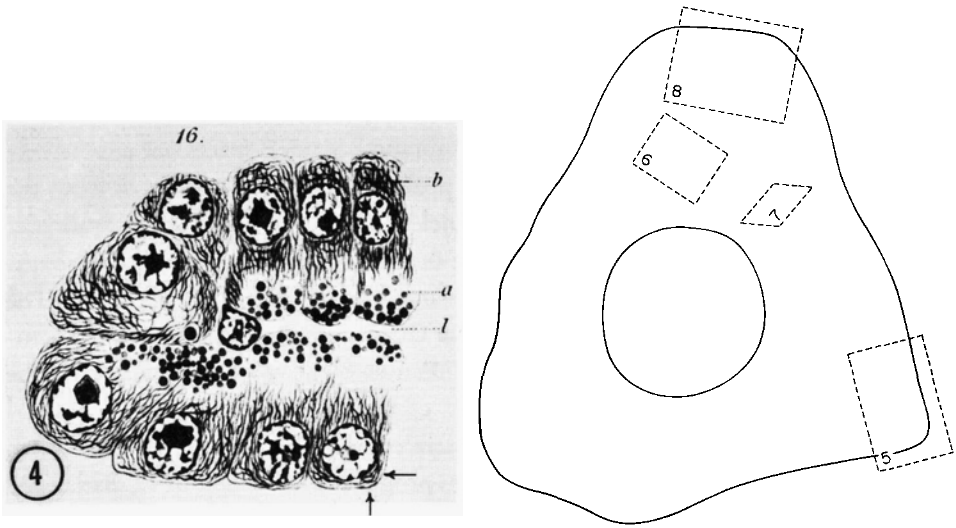
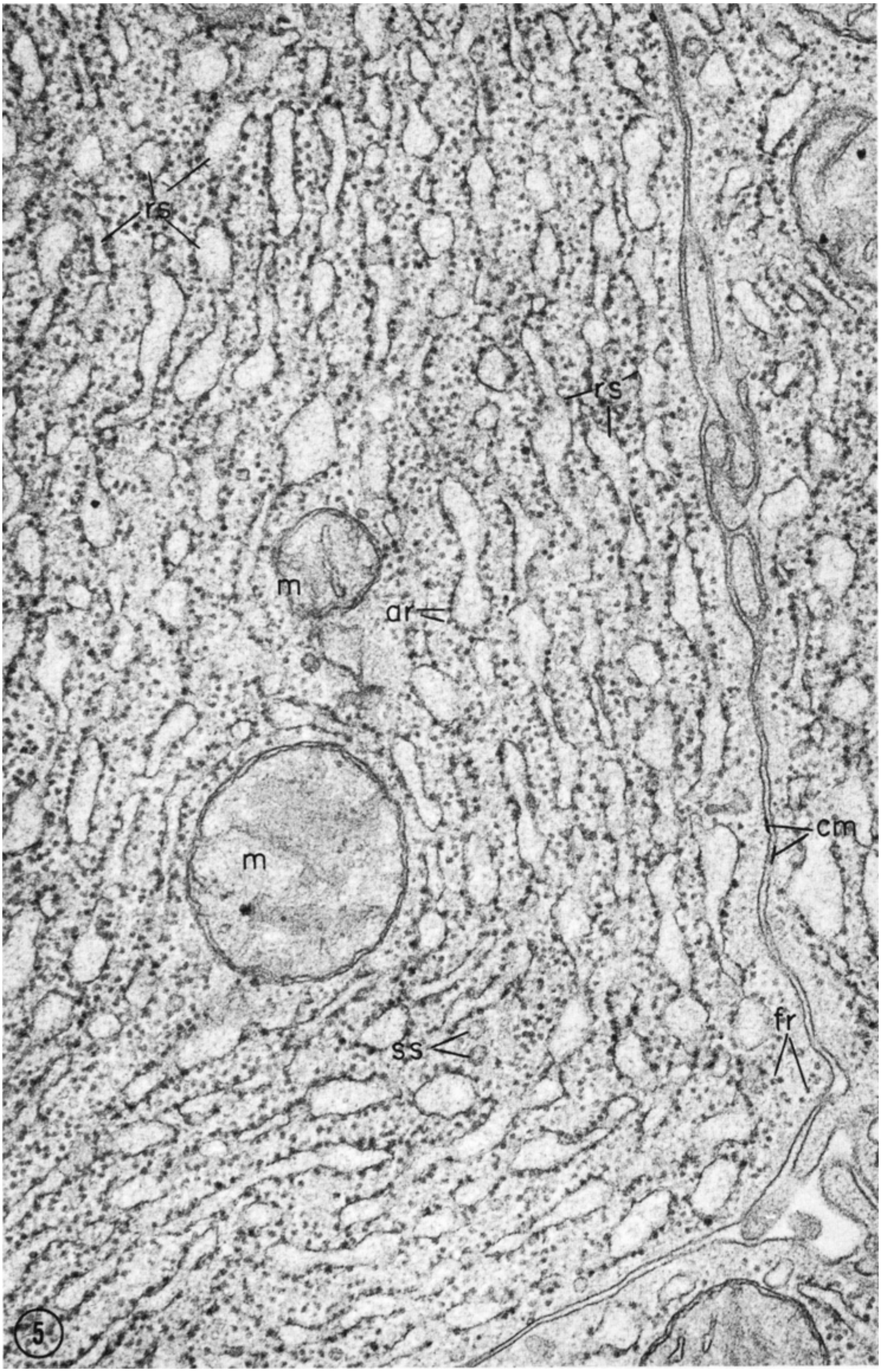


FIGURE 4. The figure shows an acinus in the frog pancreas. The basal region is occupied by a basophilic fibrillar material. The arrows (added by me) mark a region comparable to that shown in Figure 5. The accompanying diagram represents a pyramidal pancreatic exocrine cell (guinea pig) similar to the large cells at the bottom of the acinus in Mathew's drawing. The enclosed areas marked 5 to 8 indicate the position and the size of the field of Figures 5 to 8 within such a field.

FIGURE 5. Small field in the basal region of a pancreatic exocrine cell (guinea pig). Parts of two adjacent cells appear along the right margin and in the lower right corner. × 50,000.


rs
rs
m
ar
cm
m
fr
ss
5

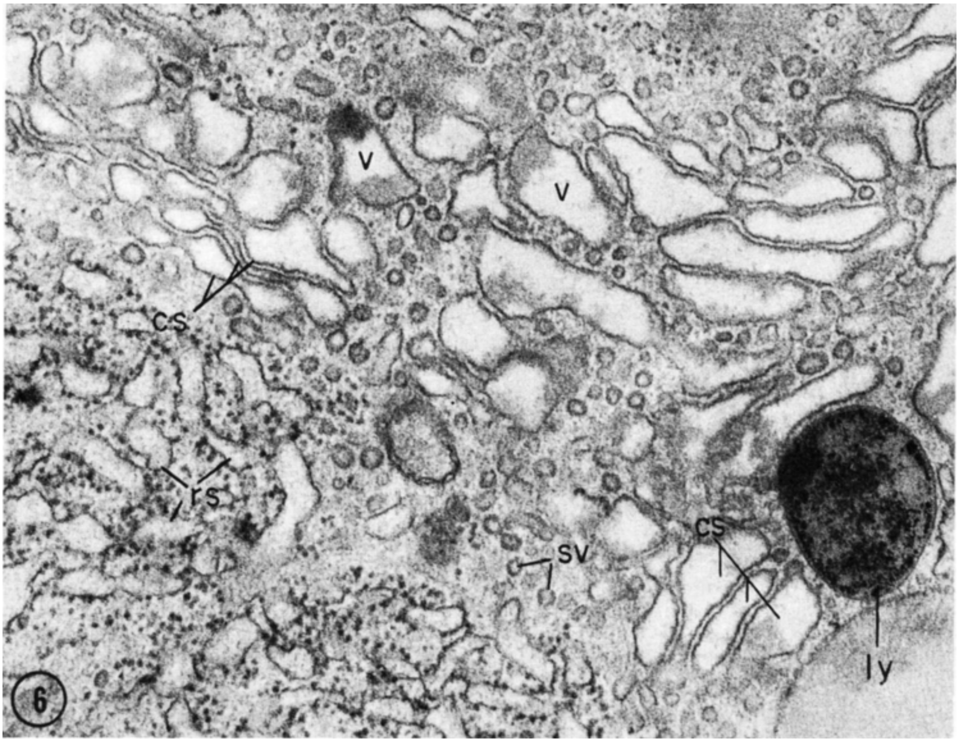


FIGURE 6. Small field in the Golgi region of a pancreatic exocrine cell (guinea pig). Rough-surfaced elements of the endoplasmic reticulum occupy the lower left corner and smooth-surfaced elements of the Golgi complex the rest of the field. The complex consists of small vesicles (*sv*), stacked cisternae (*cs*) (distended in this case to a varied extent), and large vacuoles (*v*). The body marked *ly* is a lysosome. × 50,000.

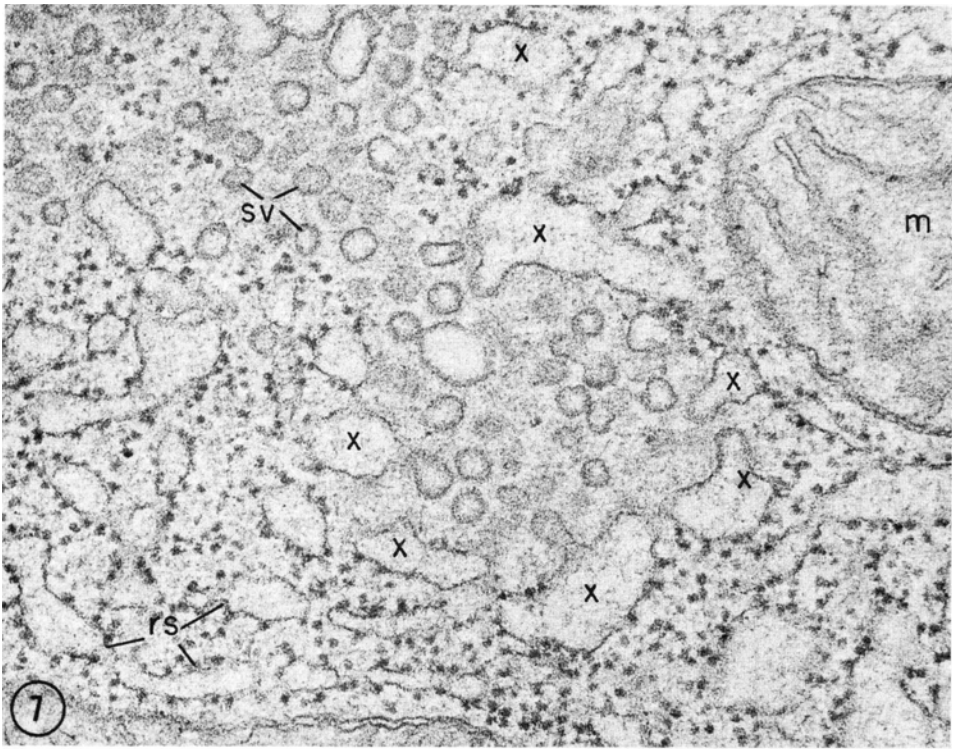


FIGURE 7. Small field at the periphery of a Golgi complex showing rough-surfaced elements (*rs*), smooth-surfaced vesicles (*sv*), and elements of intermediary appearance: part rough and part smooth. The cisternal space of the latter is marked with an *x*. × 70,000.

metazoan's intestine, and the rough part as the equivalent of a digestive gland.

A salient consequence of the existence of the endoplasmic reticulum is the partition of the cellular space into three main compartments (Figure 9). One surrounds the network, is continuous at any time, and represents the cytoplasmic space proper bounded on one side by the cell membrane, and on the other by the membrane of the endoplasmic reticulum. The other is enclosed by the limiting membrane of this system; it is usually referred to as the intracisternal space; it is occupied by ingested material or secretory products; and represents a buffer compartment between the internal and external medium of the cell. As already mentioned, the cell can operate this space as a continuum or as a succession of temporarily isolated, secondary compartments. The third major compartment is bounded by the nuclear envelope and is occupied by the nucleus.

The ribosomes populate the cytoplasmic space proper and occur either free in the cytoplasmic matrix, or attached to the membrane of the endoplasmic reticulum. Usually they are disposed in clusters when free[29] (Figure 11), or in rows when attached[24] (Figure 10). This tendency to form more or less characteristic groupings *in situ* may be related to the current hypothesis, derived from work with isolated ribosomes, that the active form of this subcellular component is a polyribosome (or polysome) rather than an isolated particle.[30] The total ribosomal population varies with the protein output of the cell; free particles predominate when proteins are synthesized for intracellular use, and attached particles when proteins are produced for export,[24] as in the glandular cell of the previous example. The attachment of the particles to the membrane of the endoplasmic reticulum is related to the already discussed functions of this system.

As in the case of mitochondria, the mass isolation of microsomes[7] (i.e., fragments of the endoplasmic reticulum) and ribosomes[24, 31] has opened a remarkably fertile field of research which has already established that the ribosomes consist of about equal amounts of RNA and protein, have a particle weight of $\backsim 4 \times 10^6$, are comprised of two unequal subunits, and represent the basic piece of cellular equipment for protein synthesis.[31] The work on ribosomes has strengthened the view that there is well-defined functional specialization among subcellular components, but at the same time has brought forward, more forcefully than before, the concept of necessary integration among various cell organs or subcellular components.

The ribosomes apparently receive the template of the protein to be synthesized

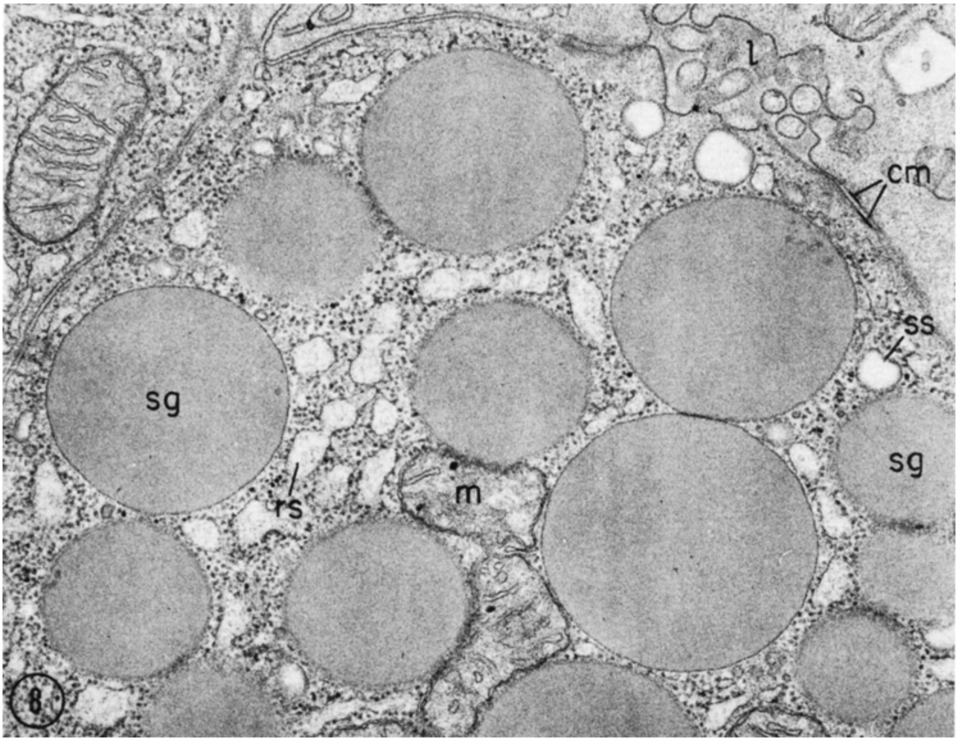


FIGURE 8. Small field in the apical region of pancreatic exocrine cell. A glandular lumen appears at *l* and parts of three adjacent cells can be seen along the upper margin of the figure. The field is occupied mainly by secretion granules, two of which are marked *sg*. × 30,000.

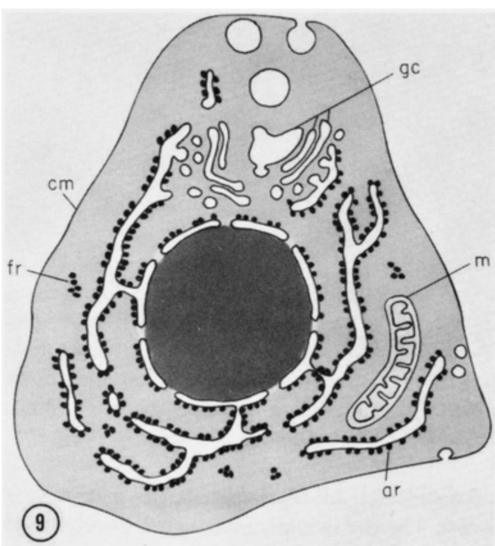
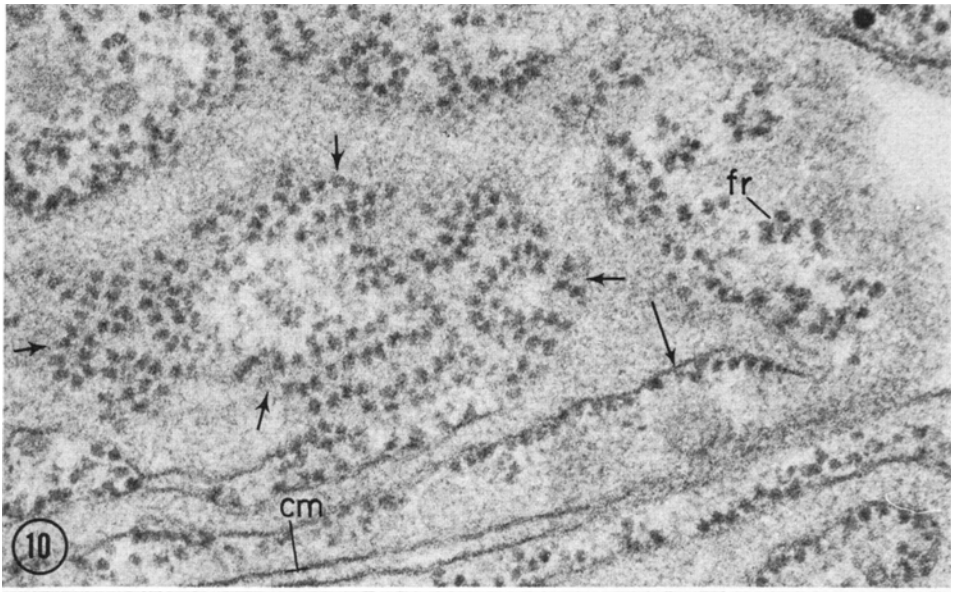


FIGURE 9. Diagram showing the three main compartments of a cell: *light grey*, intracisternal space divided in this case into a series of secondary compartments; *grey*, cytoplasmic matrix which contains as a rule a series of subcompartments (like the mitochondrion marked *m*); *deep grey*, nucleus. The external medium of the cell appears in light grey like the content of the intracisternal space. This does not mean that they are identical.



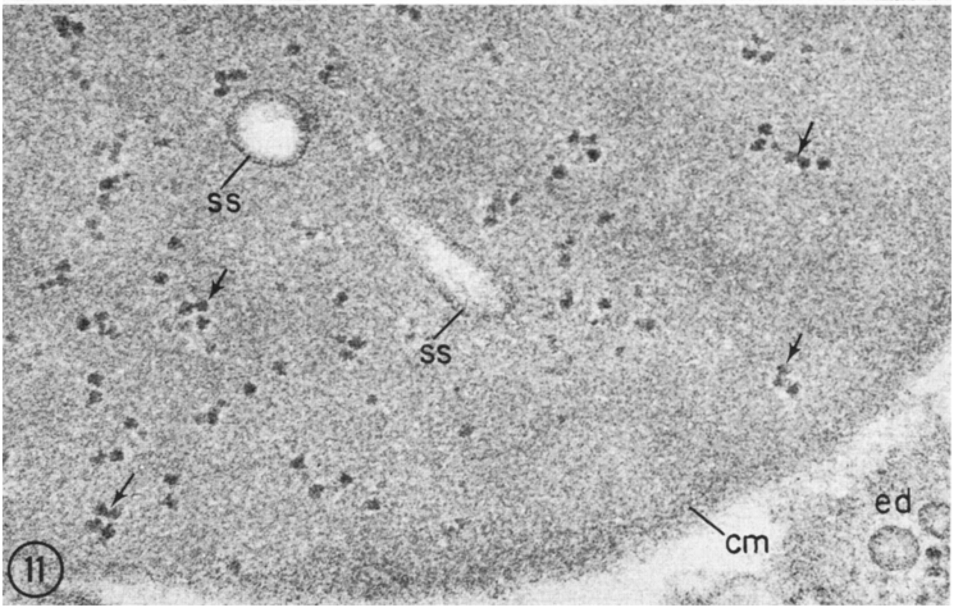


FIGURE 10. Parts of two adjacent fibroblasts (rat myocardium). Two distended, branching, rough-surfaced cisternae appear in the field. Their limiting membrane is cut normally in some places (long arrows) and obliquely in others. A grazing section, giving a full-face view of the membrane, appears between short arrows. Note the patterns (rows, double rows, and spirals) formed by the attached ribosomes on this membrane patch and in other places where the membrane is cut obliquely. Note also the clusters formed by the free ribosomes. × 85,000.

FIGURE 11. Part of a young red blood cell (rat reticulocyte). All ribosomes are free and many of them appear in small clusters or short rows (arrows). The endoplasmic reticulum is poorly developed and consists only of smooth-surfaced elements. The cell membrane obliquely cut appears as a broad, poorly defined band. *ed* marks an adjacent endothelial cell. × 85,000.

from the nucleus, more exactly from one or more cistrons in a certain chromosome in the form of a messenger-RNA molecule;[32] they accept amino acids activated by enzymes located in the cytoplasmic matrix[33] and brought to the ribosomes by transfer (soluble) RNA molecules;[34] they use a polymerase to link these amino acids by peptide bonds;[35] and finally discharge the ensuing protein either directly into the cytoplasmic matrix or across the membrane of the endoplasmic reticulum into the cisternal space, depending on the final destination of the product. The energy needed for all these operations is supplied mostly by the mitochondria. In this case, the salient aspect is the extent of integration, which is superimposed on functional specialization, and apparently involves practically all important cell organs. This implies elaborate coordination which in part may have a structural basis. The fact that the extent of supplementation, by "cell sap" components, needed for full activity varies considerably among ribosomal preparations[36] suggests that the "soluble" enzymes and carriers involved in the preliminary steps of protein synthesis are organized in functional complexes centered on ribosomes or polysomes. Presumably, most of these structures are destroyed by our preparation procedures.

Cytoplasmic matrix: In addition to ribosomes and ancillary apparatus, the cytoplasmic matrix contains, in unknown degrees of organization, most of the enzymes of the intermediary metabolism of the cell, including those involved in preliminary or alternative pathways of ATP synthesis, such as anaerobic glycolysis[37] and phosphogluconate oxidation.[38] It also accommodates the metabolic reserves of the cell, in the form of polysaccharide (glycogen) granules[39] or lipid droplets, and its aqueous phase contains the main, but not the only, cellular pool of soluble precursors. The matrix also houses a population of contractile protein molecules, varied in size and degree of organization, which is responsible for most cell movements.

Nucleus: At this point we can leave the cytoplasm, although the inventory of local structural patterns is still far from being exhausted,[40] and move into the nucleus (Figure 12) as if tracing back the path followed by messenger RNA's. They probably reach the cytoplasm through the numerous pores which interrupt the nuclear envelope,[26] but within the nucleus their pathway is practically unknown except for the starting point which is assumed to be on a chromosome, or more precisely, on a DNA molecule. Although the nucleus was the first cell component to be isolated in mass almost a century ago,[41] and although reasonably

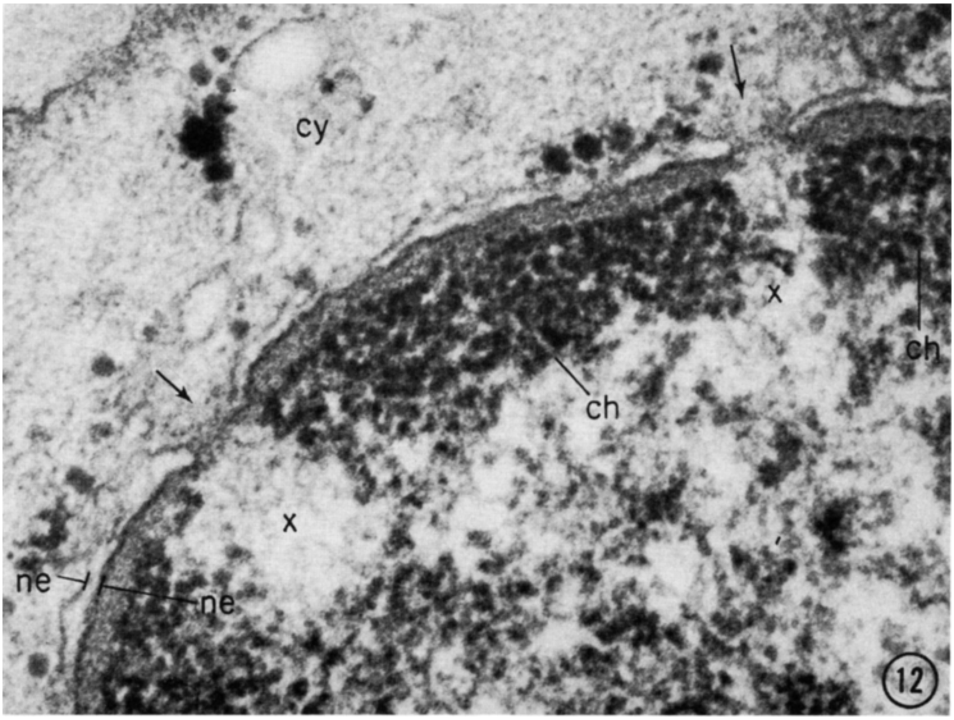


FIGURE 12. Part of the nucleus of a smooth-muscle cell (frog skin artery). The two membranes of the nuclear envelope are marked *ne*, and two nuclear pores are indicated by arrows. The dense masses of "granular" texture (*ch*) are probably chromosomes of this resting nucleus. The lighter material that fills the "channels" (*x*) leading to the nuclear pores is mostly protein. A band of dense, relatively homogeneous material separates, in this case, the chromosomes from the nuclear envelope. × 120,000.

good preparations of whole nuclei can be obtained,[42, 43] the isolation of subnuclear components, such as chromatin, chromosomes, and nucleoli, has been less successful.[43] Even the morphological analysis of the nucleus is much less advanced than that of the cytoplasm. The enzymes involved in the replication of the genetic code (DNA) appear to be located, at least in part, in the nucleus,[44] together with those which carry through the transcription of the code into messenger RNA's.[45] Moreover, there are indications that all cytoplasmic RNA's, including ribosomal and soluble RNA's are synthesized in the nucleus,[46] but the evidence in case is still questioned. Finally, the possible role of certain nuclear proteins, the histones, in the repression or activation of genes, begins to be analyzed.[47] Yet the components directly involved in all these operations probably represent a small fraction of the mass of the nucleus. The rest is still in the shadows. There is, however, substantial evidence that at least in some cell types the nucleus comprises a protein-synthesizing system similar to that of the cytoplasm.[48] This finding brings forward what appears to be still another principle of cell organization, namely the partial decentralization (in terms of cell compartments) of certain basic processes such as protein synthesis. Recent evidence suggests that autonomous protein-synthesizing systems operate not only in the nucleus but also in chloroplasts[49] and mitochondria,[50] possibly supported in part by a local autonomous DNA code.[51, 61] The situation has been regarded as vestigial symbiosis;[52] it could be just as well viewed as intracellular differentiation.

Cell membrane: The entire cell is separated from its external medium by a thin ($\backsim$80 Å) membrane which appears to be a stratified structure comprised of a bimolecular layer of polar lipids covered on both sides by protein films.[53] It controls the exchanges between the aqueous content of the cell and its equally aqueous surroundings first by interposing a practically impermeable barrier, the bimolecular lipid leaflet, and then by modifying this barrier through a variety of means envisaged at present as pores, carriers, and energy-driven pumps, all operating on individual molecules or ions.[54] The result is selective permeability which has basic features common to the generality of cells, but details characteristic to each cell type. The cell membrane is probably a mosaic of functional units whose heterogeneity is compounded by the fact that local differentiations affecting stratification, thickness,[55] and enzyme activities (presumably connected with active transport) may appear on parts of the membrane exposed to different media.[56]

The cell apparently controls not only exchanges between its external and internal medium but also exchanges among the various compartments in which its own internal medium is subdivided. The membranes limiting such compartments (endoplasmic reticulum, mitochondria, etc.) are known to be rich in polar lipids,[7] generally have a layered structure comparable to that of the cell membrane,[57] show complex enzymatic activities,[58] and at least in some cases are known to be semipermeable.[18, 59]

EVOLUTION AND DIFFERENTIATION AT THE CELLULAR LEVEL

Since the structural framework described is that of an animal cell in an advanced state of differentiation, one can ask to what extent the prototype of the framework, if such exists, is affected by this condition. The answer is given by the following examples, which represent a precipitous journey against the double current of differentiation and evolution towards more simple, supposedly earlier forms of the framework.

An undifferentiated animal cell consists of the usual combination of structural patterns. The differences are only quantitative: it has a less well-developed endoplasmic reticulum and a larger population of clustered, free ribosomes than its differentiated counterpart.[24]

An algal cell has a nucleus, mitochondria, an endoplasmic reticulum, Golgi complexes, and ribosomes[60] built along the same patterns as in animal cells (Figure 13). In addition, algal and plant cells in general contain an important organ specialized in photosynthesis, the chloroplast, which, like the mitochondrion, has a large amount of internal membranous material but shows more elaborate secondary compartmentation.[61] The organization of a fungus cell (hypha) [62] (Figure 14) is even closer to that of an undifferentiated animal cell. In a bacterial cell, however, the situation changes (Figure 15). The cell volume is about 500 times smaller than that of an animal cell. The only easily recognizable components are the cell membrane and the ribosomes, which have been extensively studied and found to be comparable in structure and function to those of animal cells. The rest is definitely different. A central irregular zone in the protoplasm is occupied by fine fibrils, probably DNA molecules.[63, 64] This nuclear zone is sometimes in contact with membrane impocketings, called mesosomes,[65] which are presumably modest precursors of mitochondria.[66] The cell has no recognizable endoplasmic reticulum, no perinuclear cisterna, hence no well-defined nucleus. Moreover, bacterial cells have no recognizable mitochondria, although it is known that they are effective in oxidative phosphorylation and that this activity is associated with a membrane fraction, presumably derived from the cell membrane[67] and possibly from mesosomes.

It seems, therefore, that in this case the rule "same function — same structure" which clearly applies for ribosomes is no longer respected. Yet if the existence and functional role of elementary particles be satisfactorily established in animal mitochondria, a new explanation becomes possible. A combination of elementary respiratory and phosphorylating particles may exist in all living systems from bacterial to animal cells, and the differences encountered may concern only a secondary superimposed structure, namely the membranous framework to which the particles are attached or in which they are embedded. Of course it remains to be seen first whether "elementary particles" can be identified at the surface of the cell membrane or of the mesosomes in bacterial cells. A similar situation may apply to the variety of extant photosynthesizing structures. An elementary unit, the "quantasome," has been recently described as being present in or on chloroplast membranes.[68] Here again the discrepancy between similar function but dissimilar structure could be explained by the existence of a common elementary particle attached to a membranous framework of increasing degrees of elaboration from photosynthesizing bacteria to blue-green algae and green plants.

Although it is known that the genetic code consists of DNA in bacterial as well as animal cells, and although chromosomes are supposed to exist in all cells, morphological studies have clearly shown that the bacterial DNA strand and the animal cell chromosome do not represent equivalent levels of organization. Here again the existence of a common elementary particle close in dimensions and chemistry to the bacterial nuclear fibrils and the addition of a series of secondary complications introduced by the folding and packing of such structures could explain the difference. Bacterial flagella and animal-cell cilia may be related in a similar manner.

The main conclusion of this survey is that there are already far-reaching distinctions among the units we call cells. Recent work has stressed the extensive and striking similarity of many subcellular patterns from fungi to mammals; it has also revealed that the structure of bacteria has little in common with that of

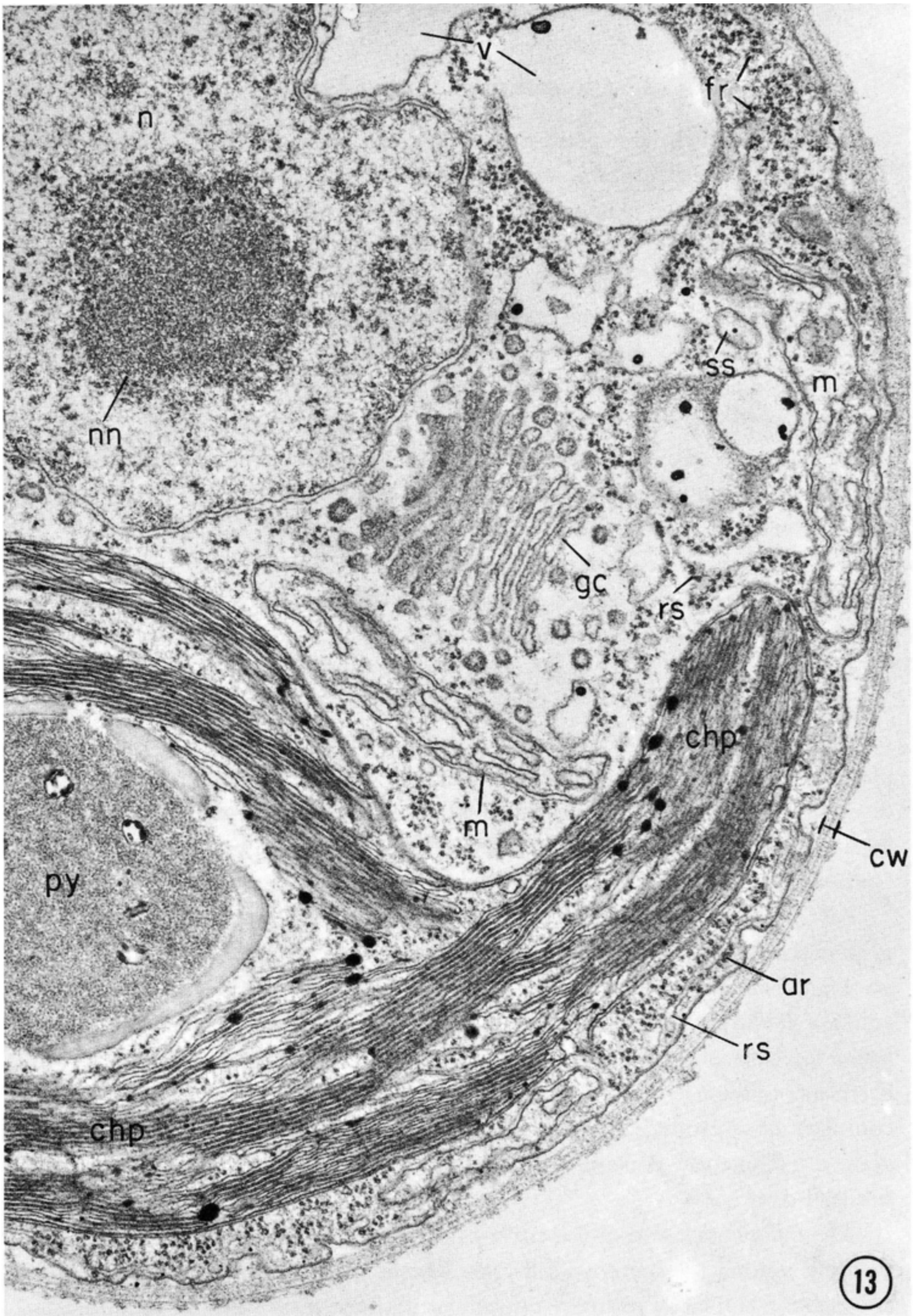


FIGURE 13. Part of an algal cell (*Chlamydomonas reinhardi*). In addition to the usual cell organs i.e., nucleus (*n*) with a large nucleolus (*nn*), mitochondria (*m*), Golgi complex (*gc*), rough- (*rs*) and smooth-surfaced (*ss*) elements of the endoplasmic reticulum, attached- (*ar*) and free (*fr*) ribosomes, the cell contains a large chloroplast (*chp*) with a pyrenoid (*py*). Vacuoles are marked *v* and the cell wall *cw*. × 25,000.

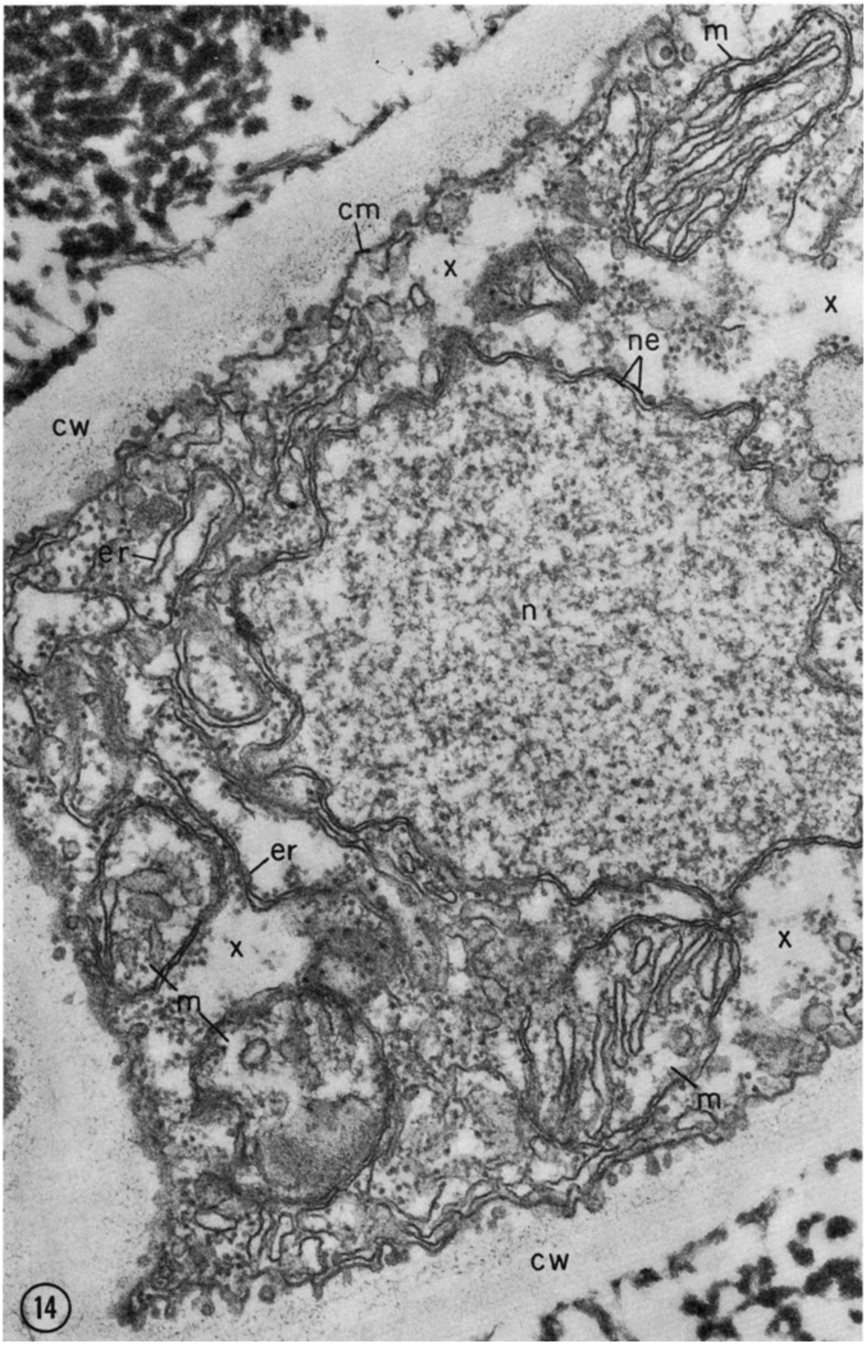


FIGURE 14. Part of a hypha of *Neurospora crassa*. *n,* nucleus; *ne,* nuclear envelope; *m,* mitochondria; *er,* elements of the endoplasmic reticulum, most of them smooth-surfaced; *cm,* cell membrane; *cw,* cell wall. Most of the ribosomes of the cell are free in the cytoplasmic matrix. The light material in regions marked *x* is polysaccharide. × 58,500.

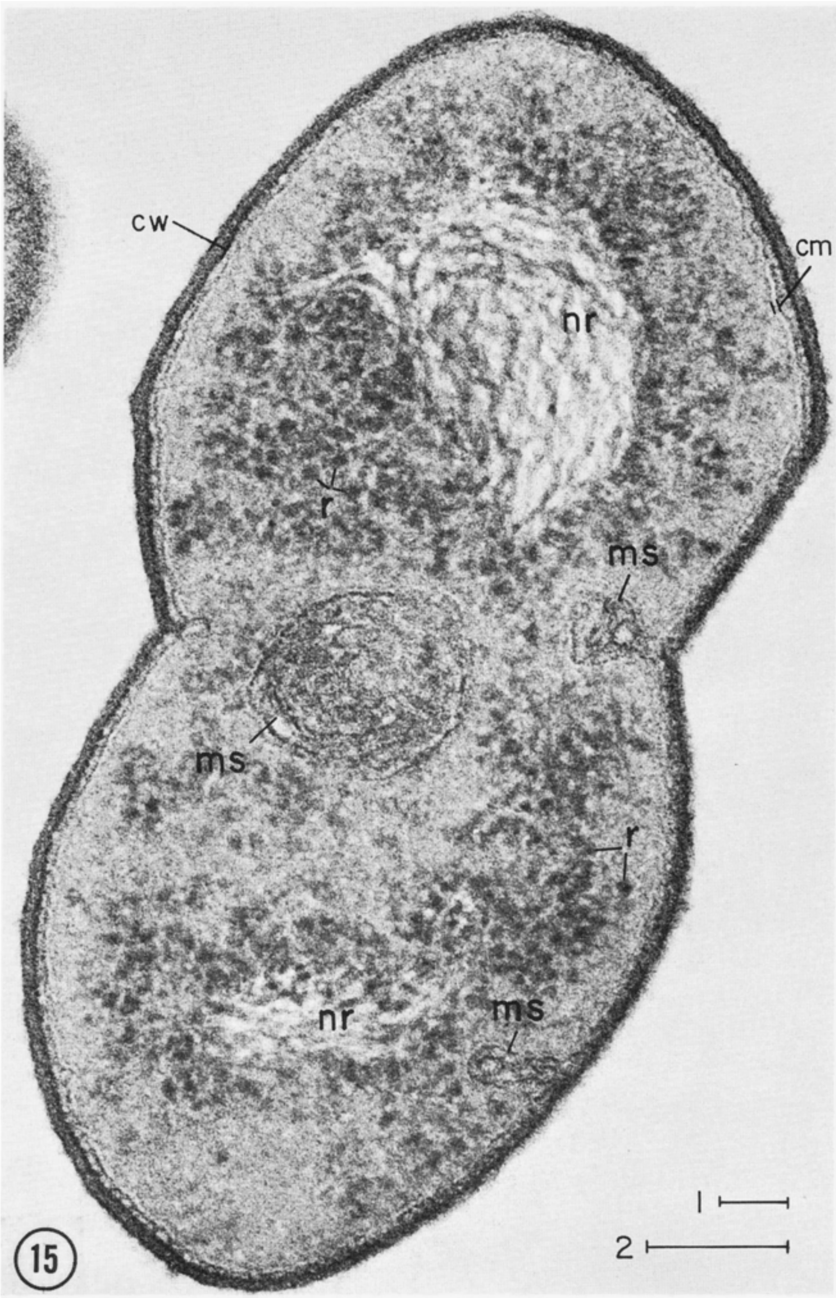
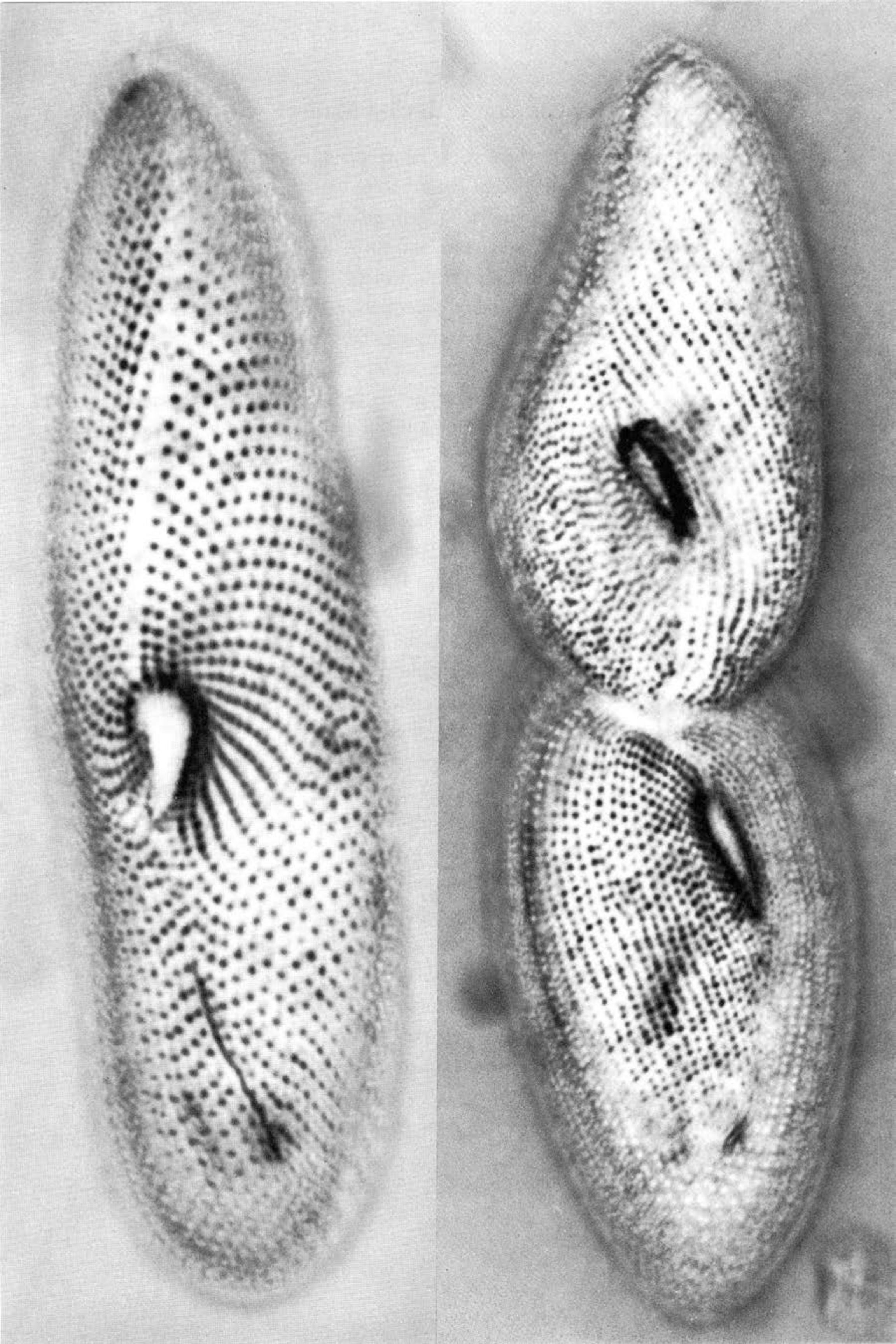


FIGURE 15. Bacterial cell (*Diplococcus pneumoniae*) in the process of division. The nuclear regions (*nr*) contain a network of fine fibrils, and are surrounded by masses of ribosomes (*r*). The protoplasm contains, in addition, three mesosomes (*ms*), two of which are in continuity with the cell membrane. The cell wall is marked *cw*. × 217,500. Marker 1 gives the transverse diameter of this cell at a magnification equal to that of the alga in Figure 13, and marker 2 at the same magnification as the pancreatic cell in Figure 5.

animal or plant cells. Clearly there is no structural unity at the cellular level. Yet it is known that at the subjacent macromolecular level of organization all cells are, generally speaking, similar. Their common functions presuppose similarly integrated multienzyme systems, and these in turn may be expected to form similarly structured organs upon their assembly. These contradictions to the widely accepted concept of the unitary organization of living matter could be explained by assuming that the various functions now supposed to be connected with definite cell organs are actually carried out by smaller elementary particles of which the ribosomes would be the first relatively well-defined example, and the "oxisomes"[69] (mitochondrial elementary particles) and "quantasomes" the next possible candidates.

The bacterial cell apparently represents the minimal but sufficient formula for the cellular level in the hierarchy of patterns of living matter. We can assume that its emergence was a crucial event in the history of life: it made replication possible, or greatly improved its efficiency; it established life and thereby triggered the evolution process. The next difficult step was probably the addition of one or two more levels of organization to the structural framework of the first simple cells leading to the emergence of the more complex animal and plant cells. The systems which succeeded in making this step apparently acquired more independence from the external medium; the possibility of using other energy sources than residual organic compounds of pre- or postbiotic origin; the ability of handling matter in mass; and the possibility of accommodating a more and more elaborate genome and of controlling its activity in time. An elaborate and controllable genome was probably the prerequisite for cell differentiation, hence for the emergence of multicellular plants and animals. Examined at the cellular level, all that followed this step appears as a rather easy exercise: a long series of variations on a common, durable, and versatile theme. It is worthwhile noting that the simple and the complex cell type stand out as sharp discontinuities in the spectrum of still extant cellular forms.

Whenever possible, I followed a historical approach in this presentation, to indicate that recent findings and present concepts are only the last approximation in a long series of similar attempts which, of course, is not ended. Time will tell how far we are at present from our final goal: the full understanding of the organization of living matter.



The Determinants and Evolution of Life

THIRD SCIENTIFIC SESSION

LEFT: Rows of granules marking the bases of the cilia in the unicellular animal *Paramecium aurelia.* The central comma-shaped space marks the position of the mouth and gullet. RIGHT: Same species showing a stage of transverse division into two daughter cells. [*preceding photograph*] 2600 times life size

Introduction

THEODOSIUS DOBZHANSKY, Chairman

MAN HAS AN abiding urge to understand himself and his place in the universe. Science is a creative response of human mind to this urge. It appears that the universe proved to be simpler and rather more easily understandable than did life or man's own nature. Physical sciences and technologies based on them, make up a lion's share of the scientific movement. And yet it really begins to look as if our time may stand in the history of science partly, if not mainly, for major advances in the study of life. Life sciences have made remarkable strides in recent years. Although there has been progress, advances of a comparable magnitude are still to come in the sciences dealing directly with man. Our bodies are "fearfully and wonderfully made," and so are the bodies of other organisms, animal and plant, down even to the simplest. Yet surely, nothing is more "fearfully and wonderfully made" than man himself.

Life sciences must deal with several, progressively more complex, levels of biological integration. The molecular, cellular, individual, populational, and biotic community levels present themselves to the observer. It is convenient to speak of molecular biology and organismic biology. In man, still another human level is superimposed, that of consciousness and culture. Some scientists cling to the reductionist faith — that the way to know everything is to concentrate on the investigation of the lowermost level, which is consequently styled the "fundamental" level. A wiser and sounder strategy of scientific research is to gain an understanding of the phenomena and regularities on all integration levels. This is not because some new and irreducible agents manifest themselves in life as compared to the inert matter, or in man as compared to the biological world, as some vitalists wanted us to believe. It is simply because every level of organization shows an integrative patterning of components from the underlying levels, and these patterns are in turn the components of the patterns on the higher, or, if you wish, less fundamental levels. Now, the patterns as well as their components equally deserve attention and study.

THEODOSIUS DOBZHANSKY The Rockefeller Institute

The contributors to this symposium will sketch for us a wide panorama of life sciences, ranging all the way from the deepest molecular to the broadest organismic level, and to the innermost workings of the human psyche. We begin, as we should, with the fundamentals. Professor E. L. Tatum will survey some of the exciting recent discoveries concerning the Genetic Determinants. Most organisms start their individual lives as fertilized egg cells, containing their sets of genetic determinants. An adult body of a human individual, for example, may contain many billions of cells, and these cells may differ greatly in structures and functions. How and why do cells become diversified during an individual's development? This is the great biological problem of differentiation. Here our guide will be Professor T. M. Sonneborn. No man is an island, and no organism is an isolated island either. It is not only embedded in its environment, but it also influences, and in part creates, its own environment and that of other neighboring organisms. This is the topic treated by Professor G. E. Hutchinson. Nothing makes sense in biology, except in the light of evolution, "sub specie evolutionis." Professor E. Mayr is the carrier of this light. And then we shall be approaching the top. For in a certain sense, all else in the life sciences is only the background for an attempt to understand man and man's mind. Professor N. E. Miller is our guide in this attempt to reach the summit.

Genetic Determinants

E. L. TATUM

IT SHOULD BE obvious that a thorough and complete discussion of genetic determinants, involving as it does, all the concepts of "molecular genetics," is impossible in the brief time available. Nor is it possible to trace in detail and chronologically the developments in this area. An attempt to do either is perhaps not even desirable in the context of this symposium. In view of the impact of molecular genetics on biological thought, often referred to as "The Revolution of Molecular

E. L. TATUM The Rockefeller Institute

Biology," it is, however, desirable to attempt to examine the important concepts involved.[1]

The importance of genes, or genetic determinants, in all of biology, needs no detailed elaboration here. It is enough to remind ourselves that their genes specify all the species and individual characteristics of every living organism, and that they determine the potentialities or limits of variability of these characters. Interaction of environment and organism can modify the expression of a character, but only within these gene-determined limits, and only as a consequence of mutation can new limits of variability be set to serve as the raw ingredients for selection and evolution.

Let us now consider the nature of genetic material, and start by listing the required properties of this material. A gene has a high degree of specificity for a given character or phenotype; it is capable of precise replication, synchronously with cell division in the case of chromosomal genes; and it segregates and recombines as a unit in sexual reproduction. Finally, it is subject to change or mutation; and the resulting mutant gene again replicates precisely and continues to function with its changed specificity.

The question of the type of molecule which would fulfill these requirements was unsolved until the transformation experiments of Avery and coworkers with pneumococcus, and the work of Hershey and Chase with bacteriophage, proved DNA, deoxyribonucleic acid, to be the basic genetic material in these forms. Much attention was consequently refocused on this substance.

Building on the findings of Pauling, Chargaff, and Wilkins, and in a stroke of "ingenius," Watson and Crick deduced and proposed the molecular structure for DNA which bears their names. This molecule consists of two strands of nucleotides, in which complementary sequences of the purine bases, adenine (A), guanine (G), and the pyrimidine bases thymine (T), and cytosine (C), are joined by deoxyribose-phosphate backbones. The two strands are held together by specific hydrogen bonding between the complementary bases A and T, and between C and G. This double helical molecule of DNA and the principle of base complementarity provide uniquely for all the required attributes of genetic material — specificity through base sequence, replication through enzymatic assembly of new complementary strands, and mutation by alteration of base sequences by addition, deletion, or substitution of bases. Through its now well established role in RNA, ribonucleic acid, synthesis, and hence in protein synthesis, DNA also fulfills the requirement of genetic material for a mechanism of specific expression.

The essential correctness of the postulated structure of DNA and its functioning in replication is supported by many lines of evidence. At the enzyme level we may cite: the synthesis by Kornberg's polymerase of DNA similar in composition, base ratios, and nearest neighbor relations, to the primer DNA serving as template; and the synthesis of the predicted type of RNA by Weiss's DNA-dependent RNA polymerase. At the molecular level we should mention the separation of the two strands of DNA and their re-formation into the double helical condition in Doty's "melting" experiments.[2]

At the chromosome level the validity of the general hypothesis is supported by the chromosome-labeling experiments of Taylor with plant cells,[3] and with bacteria, the experiments of Meselson, Stahl, and others. The results of all such experiments suggest a semiconservative mechanism of strand replication, consistent with the accepted DNA structure.

To summarize thus far, we now believe that: (1) the basic genetic material is DNA, replaced in certain viruses by RNA. (2) The information needed to specify all the species and individual characteristics of every organism is recorded in its DNA in terms of a non-overlapping triplet code of bases. (3) During gene replication this information is reproduced by the unidirectional assembly of new DNA strands, each complementary in base sequence to the single strand serving as template, with the net result being the formation of two daughter molecules of DNA, identical with the parental molecule.

Let us now consider the final step — the translation of genetic information into protein structure. This is believed to involve essentially the transcription in the nucleus of a specific DNA code sequence into the complementary sequence of bases in single-stranded RNA molecules. The demonstration by Rich that complementary DNA and RNA strands can form hybrid molecules intriguingly suggests a possible mechanism for the transcription of DNA into "messenger" RNA. This phenomenon has successfully been used by Rich, Spiegelman, and others to detect complementarity between DNA and RNA strands, and to isolate specific complementary RNA. This "messenger" RNA leaves the nucleus, and serves as the template for assembly of amino acids into a polypeptide chain. This assembly takes place in the cytoplasm, in the protein factories of the cell, the ribosomes.

Each specific DNA base triplet thus ultimately specifies a particular amino acid, and the sequence of base triplets specifies the sequence of amino acids in the final polypeptide product, the protein or enzyme.

In more detail, it appears that in this process amino acids are first activated enzymatically and attached to specific small RNA molecules, called "transfer," "adapter," or "soluble" RNA. It has been suggested that the molecule is bent in the middle and coiled back on itself in a double helix, leaving the bound amino acid at one end and three unpaired bases in the bend at the opposite end. These three bases would bind the molecule to three complementary bases in the messenger RNA. This could bring the amino acids together in the proper order and spatial relationship to form polypeptide chains in the sequence specified originally by the DNA base triplet sequence.

Some additional aspects of the replication and transcription processes deserve mention here, since they either add to our understanding of the mechanisms involved or emphasize important areas of incomplete understanding.

Present evidence suggests that in contrast to the "autocatalytic" replication of DNA, in which both strands are necessarily involved, in the "heterocatalytic" function of RNA synthesis only one DNA strand is read and transcribed.

The two processes may involve reading from different geometries of the DNA molecule. Thus, actinomycin,[4] which specifically binds to guanine residues in DNA, and which appears to rest in the minor groove of the helix, completely inhibits cellular RNA synthesis at concentrations which do not affect DNA synthesis. This constitutes substantial evidence that all cellular RNA synthesis is DNA-directed.

Significantly, actinomycin has no effect on the cellular replication of RNA viruses. The logical inference from this is that viral RNA serves as its own template, and is independent of a DNA template. This inference has recently been supported by experiments on RNA production in virus-infected, actinomycin-treated cells, and by the demonstration of the presence in virus-infected cells only of a virus-specific enzyme, RNA polymerase, involved in the replication of viral RNA.

According to present evidence, replication of an RNA virus involves the heterocatalytic functioning of this RNA as messenger RNA, leading to the synthesis in host cell ribosomes both of virus-specific enzymes including RNA polymerase and of virus proteins, and also involves the autocatalytic functioning of the viral RNA as the template for its own replication. Thus, the molecular events in RNA virus infection and multiplication are consistent with, and add to, our understanding of molecular genetics. I do not wish, however, to give the impression that our understanding of DNA replication and transcription is complete. We might mention a few problems as examples of areas of considerable uncertainty, but of great interest and activity. What is the molecular basis of punctuation in a genetic determinant — in other words, what determines where replication and transcription begin and end? The answer may well lie in the answer to another question: what are the functions, if any, of the bases in DNA other than A, T, C, and G? And, is the apparent linkage of a few amino acids in DNA of any significance in its functioning?

In addition to questions such as these, our knowledge of the details of structure and function of genes is far from complete. Whole areas remain, such as the functional and structural relations of euchromatin and heterochromatin in chromosomes, and the precise arrangement of genetic DNA in chromosomes during the various phases of nuclear and chromosome activity, division, and recombination. The bridging of this gap between structure and function at the molecular and at the cellular level is one of the exciting and challenging problems in modern biology.[3]

Another exciting and important area of molecular genetics which is advancing at a rate which could not have been predicted a few years ago is that of the "genetic code" — the identification of the base triplet sequences which specify the various amino acids in protein synthesis. This "breakthrough" was initiated by the demonstration by Nirenberg and Matthaei[5] that polyuridylic acid coded for phenylalanine in the *E. coli* system of *in vitro* protein synthesis, and was facilitated by the availability of the various mixed polymers of RNA as synthesized by the Ochoa enzyme. As the result of several years of intensive work, especially in the laboratories of Nirenberg and of Ochoa, it is now possible to assign triplet RNA base codes to all the amino acids. The corresponding DNA triplet code should then be the complement of the RNA code. Tests of this code in other systems, bacterial and animal, have indicated that the *E. coli* code is in fact universal, a finding of considerable evolutionary significance. This in turn has made it feasible to interpret the increasing numbers of known genetically determined single amino acid replacements in proteins in terms of the code, as in hemoglobin, or the enzyme tryptophan synthetase. It is already possible to assign to some amino acids not only their triplet of bases, but also the order of the bases within the triplet, by a deductive process of elimination analogous both in method and challenge to solving a crossword puzzle. Considerable progress is also being made in turning this process around and starting from amino acid differences in homologous proteins, from different species, to deduce the number of mutational steps involved in their divergent evolution. The most complete studies of this type are that of Margoliash[6] on cytochrome *c* and of Ingram on hemoglobin.[7] The soundness of this approach is supported by the success of Yanofsky[8] and coworkers in their correlation of experimental mutation and intragenic recombination with amino acid substitutions in the A-protein of tryptophan synthetase of *E. coli*. Gratifyingly, the results are consistent with the accepted code.

One more aspect of the status of the code problem should be mentioned. This is the frequence of multiple coding[9] for a number of amino acids, such as leucine. This phenomenon of redundancy is often referred to as "code degeneracy," a term appropriate to cryptography, but with undesirable biological connotations. In the case of leucine it appears that there are two different transfer-RNA molecules, each specific for one of the two alternate triplet codes for this amino acid. This redundancy may have functional and evolutionary biological advantages, not implicit in the term "degeneracy."

We have implied, but not specifically stated, that gene mutation in molecular terms consists at its simplest level of replacement of a single base in a code triplet, so as to result in the substitution of one amino acid by another in the polypeptide product. This simplified picture is consistent with estimates of the smallest size of the mutational unit, as calculated by Benzer from his studies on mutation in bacteriophage.[10] It is also consistent with the known effects of gene mutation on amino acid replacements in proteins such as hemoglobin, as shown first in the "finger-printing" experiments of Ingram, and in the enzyme studied by Yanofsky, and already referred to. In addition, this picture is consistent with mutational experiments with base analogues, which presumably are incorporated into DNA during its replication, and subsequently through faulty pairing leading to base replacement, express their effects as mutations. It is of considerable interest in this connection that the use of base analogues with their specificity of action in inducing base substitutions in DNA and in correcting phenotypic expressions of mutant DNA by replacing bases in RNA, has enabled Benzer to deduce a fairly complete base sequence map of his RII locus of bacteriophage. Thus, recent studies in the area of genetic fine-structure, whether from the standpoint of structure, mutation, function, or recombination, are so far consistent with the concepts of molecular genetics.

Even the phenomenon of "complementation"[11] which at one time appeared inconsistent with the implicit relation of gene and protein, is now viewed as falling into this consistent pattern. This phenomenon, in which two allelic mutant genes, either in different chromosome strands or in different nuclei, join in controlling the nature of a single enzyme, is now believed to be due to the assembly of several finished polypeptide subunits into the functional protein.

Another example of our improved understanding of gene interaction relates to the functioning of suppressor genes. Restoration of normal function to a mutant gene by a "suppressor" mutation at another locus is a fairly general phenomenon, difficult to interpret within the general framework of our concepts of gene structure and function. It has recently been suggested by Yanofsky that some suppressor genes affect the amino acid specificity and affinity of a transfer RNA such that it transcribes a mutant triplet in messenger RNA into the amino acid specified by the original triplet.[8] This attractive hypothesis is indeed supported by amino acid replacement studies, and by the "pleiotrophic" effects of certain "universal" suppressors in restoring the normal functioning of several different, unrelated mutant genes.

Although the concepts of gene structure and function so far discussed are very satisfying, they leave unexplained a major problem of biology, the functioning of genes in development and differentiation. Although Dr. Sonneborn is discussing this topic next, certain aspects bear directly on the nature and functioning of genetic determinants, and should be mentioned at this time.

If, as is generally accepted, the genetic complement of cells does not change during their differentiation, gene activity must be regulated during this process. This implies interactions both between genes, and between genes and intracellular and extracellular environments so as to result in an orderly sequence of stimulation and repression of gene activity. That such differential activities do indeed occur during development is supported by cytological and histochemical evidence on "puff" formation and RNA production by specific chromosome bands in amphibians and insects, and is consistent with the phenomenon of nuclear differentiation demonstrated by Briggs and King in amphibian embryos. Interactions between genes and their immediate environments are known at the genetic level as in the classical "position effect" in which the quantitative action of a gene varies with its location on the chromosome in relation to other genes and to intragenic "heterochromatin." At the molecular level, RNA synthesis varies in relation to the type of histone associated with the DNA, as shown by Bonner and by Mirsky and Allfrey.

Conceptually, it is satisfying that there appear to be two major classes of chromosomal genes, one being structural genes, determining the quality of proteins, and the other consisting of regulator genes, controlling the time and quantitative rate of action of structural genes. In the model system, involving the synthesis of the *E. coli* enzyme β-galactosidase, Monod, Jacob, and coworkers[12] have provided evidence, primarily genetic, that the functions of several linked genes within a limited region of the chromosome, the operon, are turned on and off simultaneously by a single adjacent gene, the operator. The activity of this operator is in turn controlled by a regulator gene, the "inducer" or "repressor" action of which is responsive to intracellular levels of substances which are substrates or products of the reaction controlled by the enzymes of the structural genes within the operon. The general occurrence of this type of genetic control system has not yet been established, but its potential significance in development and differentiation is obvious. Certainly, control of enzyme synthesis through induction and repression by substrates and end-products, and the inhibition of enzyme activity by end-product feed-back mechanisms, are known in higher organisms as well as microorganisms.

This bird's-eye view of genetic determinants would not be complete without mention of non-chromosomal determinants.[13] A non-chromosomal determinant can be detected only if its presence and absence are correlated with a phenotypic difference which does not follow a Mendelian pattern. In most cases it follows a pattern of complete maternal inheritance, and does not segregate in genetic crosses. To add to the difficulty of detection, non-chromosomal determinants are usually present in a cell in high multiplicity, and their multiplication is not usually linked to cell and nuclear division. Because of these characteristics it is not surprising that relatively few examples are known, primarily in microorganisms.[13] The best-studied examples include chloroplast inheritance in plants, the "petite" character in yeast studied intensively by Ephrussi, the similar "poky" character in *Neurospora* found by the Mitchells, the streptomycin resistance characters discovered in *Chlamydomonas* by Sager, and the drug-resistant characters studied in bacteria by Watanabe. Not all non-chromosomally or maternally inherited characters properly belong in this category. In a number of instances non-chromosomal inheritance has eventually been found to involve endosymbionts or parasites such as several viruses in *Drosophila* and the Rickettsia-like entities in *Paramecium* and in other protozoa. Actually, the distinction between a non-chromosomal but truly cellular entity and an "endosymbiont" becomes almost purely semantic in certain instances. This becomes particularly evident when one considers that a certain class of genetic determinants, called "episomes," can exist and reproduce either in direct association with chromosomes or independently, and can transmit chromosomal genes from one cell to another.[14] The bacterial virus, *lambda*, for example, in common with other temperate viruses, may replicate in synchrony and harmony with the bacterial chromosome on which it occupies a specific site, and can also incorporate and transfer small segments of host genetic material from cell to cell as in transduction. *Hfr*, as a gene on the bacterial chromosome, determines a high frequency of gene recombination — hence its name — and as the free F^+ factor, transforms F^- (female) bacterial cells to F^+ (male) cells, in addition to transferring host genes in "sex-duction" or "F-duction." Regardless of the precise definition and nature of the various entities involved in non-chromosomal inheritance, all apparently contain DNA, and in the same sense as is a gene, are "self-reproducing." It is also clear that non-chromosomal or episomal inheritance are phenomena which must be considered in relation to development.

In conclusion, although our knowledge of the nature and functioning of genetic determinants has been increasing exponentially over the past decade or so, and we are approaching ever closer to a real understanding of the molecular basis of life and heredity, even more remains to be learned.

Within the next hundred years great advances can be expected in the control of mutational processes, in the design and synthesis of genetic determinants, and in the development of techniques for the introduction of such new genetic determinants into the genome of living organisms. The next centennial program of this Academy may very well include a symposium on "Genetic Engineering and Controlled Evolution." With the increase of our knowledge in these and other areas, it is sincerely to be hoped that our ability to use this knowledge wisely will increase in proportion.

The Differentiation of Cells

T. M. SONNEBORN

INTRODUCTION

THE DIFFERENTIATION OF CELLS, as I shall try to make clear in a moment, is one of the most fundamental and fascinating problems of biology. After a long history of ups and downs, extending over a period longer than the century of this Academy's existence, attempts to solve the problem appear now to be making remarkable progress. There is widespread and growing conviction among biologists that this will be the area of one of the next great triumphs of biology in the decades immediately ahead. As you will see, the progress already made provides some basis for this optimism. But first I shall explain the nature of the problem and of the current reorientation of research.

THE PROBLEM

A cell consists of a nucleus surrounded by cytoplasm. The membrane-bounded nucleus contains, among other things, the chromosomes with their genes. The cytoplasm is also membrane-bounded and contains an ordered array of fluids, gels, granules, fibers, and membranes grouped into characteristic organelles.

Although life must have arisen and long evolved in forms simpler than a cell, every form of life now existing — with minor exceptions — is either a cell or, after having been a single cell at one stage of its life, came to be composed of many cells. The cell is the basic unit of structure and function in existing organisms.

The differentiation of cells is most familiarly associated with cell division. We are all aware, for example, that a human being starts life as a single cell, a fertilized egg. The egg divides into two cells, then each of these into two more, and so on through many successive cell divisions until the more than 10^{15} cellular

T. M. SONNEBORN Department of Zoology, Indiana University, Bloomington, Indiana

building blocks of the human body have been formed. Even this is not the end, for cell divisions continue in certain organs and tissues throughout life. During development, the first cells to arise seem to be identical, but soon diversities appear. At first they are relatively slight and generalized, but later the cellular differences become greater and greater, yielding more and more specialized cell types. For example, nerve cells are specialized for transmission of signals, and muscle cells for contraction. Finally about 100 different kinds of normal cells can be distinguished by their structure and function. Unfortunately, sometimes abnormal cells, such as tumor cells, also arise sooner or later. By differentiation of cells is meant in part this appearance of diversity, both normal and abnormal, among the cell progeny of the egg cell during the whole of the life of the individual. How this increase of heterogeneity occurs, and how its marvelously precise and regular ordering in space and time is governed so as to yield at every stage an integrated functioning multicellular individual, are among the deepest and most challenging problems of biology.

In this most familiar example, cellular differentiation is associated with cell division and so might be imagined to depend ultimately upon the parceling out of diverse parts of one cell into different daughter cells. But it is important to recognize that cellular differentiation can also occur in the absence of cell division. This happens for example as cells grow older. Cellular changes also develop in response to changes in the milieu, including contact with another kind of cell or with its diffusing or circulating products such as hormones. Transformation of cells without division is dramatically evident in the life of a slime mold such as *Dictyostelium* which has long fascinated investigators. After a colony of separate creeping amoebae, all of which may be descended from a single ancestral amoeba by repeated divisions, has exhausted the available food supply (bacteria), the dispersed cells suddenly aggregate into a mass or slug. Without further cell division, the apparently identical amoebae of the slug pile up to form an upright stalk crowned by a capsule containing spores. Some amoebae have differentiated into stalk cells, others into spores — very different from each other and from the amoebae of the slug.

Such integrated diversification without cell division among the identical cells of a multicellular unit is paralleled by the integrated diversification among the parts of a single cell. The unicellular animals include perhaps the most complexly organized of all cells. Yet some of them encyst in a very much simplified form and, when they excyst, recreate their amazingly complex but precise pattern of intracellular differentiation in an elegantly ordered series of developmental processes. Comparable precision and elegance mark the structural developments accompanying cell division. Two identical asymmetric patterns of complex structure are created from one without dedifferentiation, that is, while maintaining all the differentiated parts of the original cell (see Figure 2). I shall come back to this highly sophisticated intracellular biological engineering later for it reveals a neglected principle of cellular differentiation. At this point I wish only to stress that differentiation among cells is closely paralleled by differentiation within a cell. This parallelism and its important implications are too often ignored.

From whatever level we look at cell differentiation, we see a continuously graded series from the most stable, apparently irreversible, to the most transient. Although some students restrict the term differentiation to the extreme class of apparently irreversible changes, I have shown elsewhere[1] that there is no justification in nature for so narrow a view. Exactly the same cellular traits appear as irreversible or unalterable in some cells and as readily reversible or alterable in other closely related cells. It is not the change which varies in such cases, but associated mechanisms which result in its fixation or modifiability. The broad inclusive domain of cell differentiation then becomes virtually coextensive with cell physiology; it is a universal and all but continuous process in living nature.

Could such varied phenomena be explained on a single principle? If so, what is it? If not, are there only a few or many underlying principles? What are they? I believe already available knowledge justifies concluding that at least two basic principles are involved and I shall discuss them. However, a single hypothesis is in fact guiding most of the current investigations and there is a determined effort to see how far one can go with this alone. The results already are spectacular. The dominant faith in this fruitful hypothesis stems from a recent reorientation of thought and research which will now be set forth.

THE CURRENT REORIENTATION

The current reorientation is really quite simple, involving only a change in the basic assumption of the role played by the genes. Formerly it was assumed that the whole set of genes was active in every cell. Hence, cells that have the same set of genes cannot become diverse by reason of direct genic action. There was indeed every reason to believe, and none to disbelieve, that (with minor and negligible exceptions) cells arising by ordinary division from a common ancestral cell had exactly the same set of genes. In the process of cell division, each chromosome replicates exactly and one of the two identical daughter chromosomes passes to each daughter cell. Thus all the cells of the body, descended from the egg cell, would have the same set of chromosomes and genes. The cell differentiations arising during development therefore appear not to be due to possession of different genes. Nor, on the assumption that the genes were all performing the same primary actions in all these diverse cells, could the differentiations be due to direct genic actions. This view seemed to be reinforced by the occurrence of differentiation within single cells in the absence of cell division, allowing no possibility of change in the set of genes. Hence, it seemed useless to consider direct genic action as relevant to the problem of cell differentiation. So, attention was directed away from the genes to the cytoplasm and the cellular milieu. The milieu was often clearly changing and the cytoplasm obviously divided unequally in some cell divisions. These then became the focus of attention.

Yet there was a justifiable feeling of uneasiness. For the whole of development, including its many progressive steps of cell differentiation, was surely hereditary. And, in well-studied organisms like the fruit fly or corn, virtually every step in development was shown to be blockable or modifiable by genic mutations.

This paradox or dilemma was formally resolved by stressing correctly that what happens in a cell depends both on what genes are present and what cytoplasmic substrates are present for them to act upon. Change of either the genes or the cytoplasm changes the results. Normally, during development the genes remain the same, but the cytoplasm varies. Hence, it was argued, the cytoplasm, not the genes, is the decisive differential in cellular differentiation and the genes may be safely ignored. This view long dominated thought and research in this field.

A very different view began to emerge about 25 years ago. One by one there began to accumulate evidences for active and inactive states of chromosomal material and for cells with identical sets of genes exhibiting "hereditary" differences in gene-controlled traits. For example, such indications of active and inactive genic states, some persistent and some transient, emerged repeatedly in my studies and those of my associates on the unicellular animal *Paramecium* beginning as far back as 1937 and coming strongly to the fore by 1948 with our investigations of the genetics of certain protein antigens.[1] Without attempting to give a full history of the course taken by this change of viewpoint, we may note at once that the most deeply analyzed and most influential study, which has become a classic of modern biology, is the one on genetic control of the production of the enzyme β-galactosidase in the bacterium *Escherichia coli*. Begun by Monod and brilliantly pursued by him, his collaborators — especially Jacob — and by many others, this great series of investigations has provided and continues to provide model systems of the control of genic action.[2] As a result it is now quite clear that a gene can be responsive to signals that regulate its degree of activity from complete inactivity, or almost so, to maximal activity. Of course, it is now known that activity means primarily production of complementary RNA (so-called genic messenger) and, through the messenger, of specific polypeptide.

These discoveries and models brought the genes back to the attack on the problem of cellular differentiation. Cellular differences, even among cells with identical sets of genes, could be due to the activity of different genes in those identical sets of genes. One no longer had to think of all the genes being active all the time. This is the essential reorientation of thought which guides most current research on cellular differentiation. It provides the now dominant hypothesis — one might almost say the principle or article of faith — that cellular differentiation is ultimately traceable to and due to variable gene activity.

THE HYPOTHESIS OF VARIABLE GENIC ACTIVITY

When I reviewed[1] four years ago for this Academy the status of the hypothesis of variable genic activity as a basis of cell differentiation, it was necessary to cite evidence mainly from work on microorganisms and to argue for its generality largely on faith. Now, however, work on higher organisms has made such progress that the faith has been vindicated as the following account of representative examples will show.

One way to estimate genic activity is to examine the minute intracellular particles, the ribosomes, which are its sites. When genes are active, that is, when they are making their RNA messengers, the messengers become associated with ribosomes and the complex operates in polypeptide production. After extraction from cells, ribosomes which are not so engaged will become engaged if offered synthetic RNA under appropriate conditions. This has been used as a measure of genic activity. The degree to which extracted ribosomes will make polypeptides outside the cell when given synthetic RNA is held to measure the degree to which they are *not* already so engaged inside the cell, that is, the degree to which the cell's genes were not active in making messenger RNA.

A number of embryologists have applied this sort of test at various stages in the development of embryos. For example, Nemer[3] has done so with the sea urchin. During the early cell divisions of the egg, before cell differentiations are determined, the ribosomes appear to be relatively inactive in polypeptide production. They are, however, perfectly competent to act for when they are removed from the egg and provided with synthetic RNA, they make the expected polypeptides. Later, just before the first differentiations of the cells begin to be determined, and thereafter, remarkable progressive changes appear in ribosomal activity. When extracted from the cells and offered exogenous synthetic RNA, they form decreasing amounts of the corresponding polypeptide. The ribosomes appear to have become programmed with the organism's own gene-produced RNA messengers. In other words, before cell differentiation becomes determined, the genes are largely silent; later they become much more active. This indicates that genes of higher organisms do indeed exist in states of varying activity and that cell differentiation is correlated with changing genic activity. These indications are confirmed by other studies which, as will appear, deepen or extend our understanding.

Obviously the analysis would be greatly facilitated and strengthened if it were possible to look at a chromosome and see whether it or definite parts of it, i.e., some of its genes, were actually active or inactive. Surprisingly enough, it now looks as if this is indeed possible in favorable cases. In them chromosome parts or sometimes nearly whole chromosomes turn out to be highly condensed or tightly wound up. Genic activity is correlated with extension, with unwinding of the chromosomal thread. One striking and widely known recent example has grown out of the work of Barr and Mary Lyon.[4] It concerns the sex or X chromosome in female mammals. The female normally has two, the male one, in each cell. In certain cells of the female most of one of the X chromosomes becomes tightly coiled up at a certain stage of development and remains so thereafter in all descendants of such cells. Which of the two coils up and which does not is a matter of chance; different ones do in different cell patches of the same female. Correspondingly when the two X chromosomes of a female bear different allelic genes, only one is expressed in some cells, only the other one in other cells. Such females are for these traits a patchwork or mosaic of differently differentiated cells. The suppression of the action of one gene is correlated with the tightly coiled condensed state of one X chromosome. The expression of the other gene is correlated with the uncoiled extended condition of the other chromosome. The active and inactive states of most genes of the X chromosomes are thus visibly evident.

Equally striking and convincing visible evidences have lately been adduced for the localized activity of small regions of a chromosome, indeed for a single gene in the most fully analyzed case, while neighboring regions and their genes remain inactive. This work has been made possible by the existence of two kinds of extraordinarily large chromosomes (Figure 1). One kind occurs in the oocyte or pre-egg cells of amphibia. These so-called lampbrush chromosomes are characterized by paired lateral loops of various sizes and shapes alternating in definite linear sequence with intervening sequences of granules of varying sizes composing a dense axial strand. Gall and Callan[5] showed that RNA synthesis

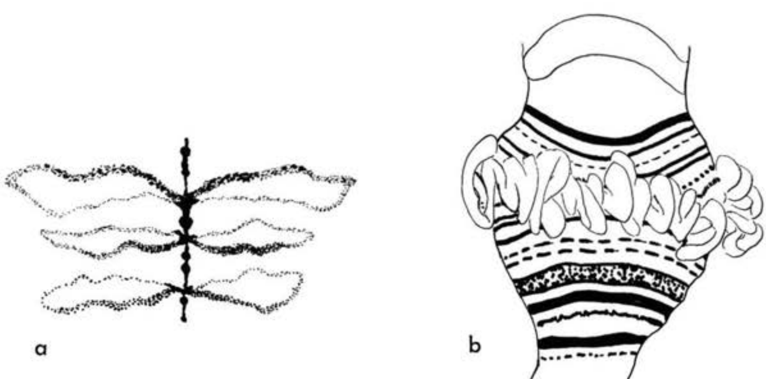


FIGURE 1*a*. Short section of a giant lampbrush chromosome from an amphibian oocyte showing lateral loops extending from axial strand marked by sequence of granules.

FIGURE 1*b*. Short section of a giant chromosome from cell of an insect showing linear sequence of bands with one of them opened up into a puff.

occurs on the loops, not on the dense axial strand. Their evidence indicates that the loops are formed by uncoiling of granules on the axial strand, and that axial granules are formed by tight coiling of the loops. The loops thus appear to be active genes, the granules inactive genes. Recently Izawa, Allfrey, and Mirsky,[6] following up earlier evidences that histones are associated with genic inactivity, showed that loops can be made to cease forming RNA and to regress to the condensed granular state by adding arginine-rich (but not lysine-rich) histone or by adding actinomycin D which blocks formation of RNA by DNA. Chemistry and the eye have again reinforced each other. Localized chromosomal sites, presumably individual genes or clusters of them, can exist in two reversible states related to the kind of polypeptide present at the site. The active RNA-forming state is uncoiled and spun out; the inactive state, during which RNA is not made, is condensed and associated with arginine-rich histone.

In most essential respects, comparable results have been obtained by Beermann[7] and others with the other kind of large chromosome found in the cells of certain tissues in the larvae of some insects such as the fruit fly and *Chironomus*. These chromosomes are really bundles of hundreds or thousands of identical stretched chromosomal threads arranged side by side in register. This association in effect magnifies the detailed structure of the chromosomes by two or three orders of magnitude and reveals a species-specific normal linear sequence of bands, varying in width, form, and structure, and of thickenings of varying degree called Balbiani rings or puffs. Comparative studies revealed that a given position in a chromosome sometimes appeared as a puff and sometimes as a band. In fact, a puff is an opened-up band, opened up into what seems to be a looped thread. These are reversible changes. The puffs of these giant chromosomes are like the loops of lampbrush chromosomes; the bands are like the granules in the axial strand. The parallel is chemical as well as morphological. The puffs are sites of RNA synthesis and are free of arginine-rich histone; the bands possess arginine-rich histone and do not detectably synthesize RNA. In both kinds of chromosomes localized chemical activity and inactivity are rendered visible. But in these giant chromosomes the relations to the genes are much better known. Many genes of the fruit fly were long ago mapped on them, the positions of the genes being defined in relation to the visible chromosomal bands. In some cases a particular gene was closely correlated with a particular band or puff. So in view of the chemical studies, we may regard the gene as inactive when its position in the chromosome is occupied by a band, as active when that position appears as a puff. These facts already show that a given gene is sometimes active, sometimes inactive.

Now it should be possible to ask and answer the question of whether differential genic activity is correlated with cell differentiation. And Beermann has just about done this. He found that only about 10 per cent of the bands were in the puffed, active form in any one cell at any one time, about 90 per cent being in the *in*active condition. Even more important, he observed that different bands or genes were puffed (and therefore active) in different kinds of cells and in different stages of development. Impressive as such evidence is, the most decisive evidence came from comprehensive study of one particular chromosomal spot. This spot was puffed only in certain salivary gland cells; in all others it appeared as a band. Moreover, the cells in which it was puffed, but no others, formed in abundance a distinctive substance, a granular secretion. That the formation of this substance depended on a single gene located at the spot where this puff appeared in the secretory cells was shown by crosses to a non-secreting strain and by cytogenetic studies which localized the gene at exactly that spot in the chromosome. In the non-secreting strain, the spot is occupied by a band which fails to puff in the cells that correspond to the normal secretory cells but fail to secrete in this strain.[8] This elegant one-gene one-band analysis shows clearly that activity of the secretion-determining gene is correlated with its puffed appearance, while its inactivity is correlated with its appearance as a condensed band; and that the activity of this gene is correlated with the differentiation of a cell into a secretory cell.

Of course, we would now like very much to know what makes this gene become active and how it happens that it becomes active only in these particular cells and just at a definite stage of development. Very little is known about this sort of thing in any organism. But some promising starts have come recently from Beermann's laboratory. For example, Clever[9] has noted that a definite pattern of localized puffs appears in a definite time sequence following administration of the molting hormone, ecdysone. It is as if the hormone, directly or indirectly, activated or derepressed a few genes and as if the actions of these genes derepressed other genes, and so on in a definite sequential series. This at least suggests a plausible partial answer to our questions. The derepression of one gene may be dependent upon the prior derepression of another gene. Hence, whatever stimulus derepresses the secretor gene may be able to do so only in cells in which certain other particular genes are derepressed. In this way, although the immediate stimulus might be a circulating hormone to which all cells are exposed, only those cells possessing the appropriate pattern of other derepressed genes will respond by the derepression of the secretor gene. This, of course, pushes back the problem to accounting for the origin in the first place of different patterns of derepressed genes in different cells. One clue to this has come from other aspects of Clever's studies. He finds that mere variations in the amount of a hormone, or in the relative amounts of two hormones, reaching a cell can bring about different puff patterns. Hopes are thus high that such studies will step by step fill in the large remaining gaps in our understanding of how progressive changes in differential genic activity can bring about the precise sequence of cellular differentiations characteristic of normal development.

Up to this point, the examples cited have correlated cell differences with the activity of different genes as if genes existed in only two alternative states, active and inactive. Quite aside from the technically difficult problem of ascertaining whether inactivity is total or just so low a level of activity as to escape detection, there is excellent evidence that the same gene can be active to varying degrees and that such quantitative differences in genic activity are also important in cellular differentiation. A beautiful example of this is found in the work of Markert[10] and others on the enzyme lactate dehydrogenase.

This enzyme, like a number of others, exists in a group of somewhat different forms constituting what is called a set of isozymes. In the case of lactate dehydrogenase, there are five isozymes and any cell that has one usually has all. But the cells of different tissues or organs or the same tissue or organ at different stages of development have different proportions of the five isozymes, sometimes very different. These differences have been shown to be due to differences in the relative activities of two genes. The two genes make two different polypeptides which are the building blocks of the enzyme. Each enzyme molecule is a tetramer, consisting of four polypeptides. The four may all be one kind of polypeptide or all the other kind or any one of the three possibilities of combining the two kinds of polypeptides in groups of four; i.e., 3:1, 2:2, or 1:3. The relative amounts of the five isozymes formed when the two kinds of polypeptides are mixed in varying proportions *in vitro* are exactly what would be expected by chance combinations. The same chance relative amounts are found in different kinds of cells, indicating that combinations are also random in the cells and depend upon the relative available amounts of the two kinds of component polypeptides. In other words, since the polypeptides are genic products, the different proportions of the isozymes found in different kinds of cells appear to be due to relative differences in activities of the two genes involved.

In sum, it is now abundantly clear that in higher organisms as in microorganisms genes may be turned off or on and turned on to varying degrees, and that such variations in genic activity result in differences among cells that have the same set of genes. The mechanisms controlling genic activity are just beginning to be explored experimentally with some success. On this phase, theory is far ahead of knowledge, so that while the specific mechanisms are largely unknown, in principle it is easy to see how controlled variable genic activity could explain many well-known problems of cell differentiation. For example, one can now readily imagine why the cells that give rise to red blood cells differentiate so as to make hemoglobin, why they make different kinds of hemoglobin in different stages of development, why a liver cell differentiates to make glucose-6-phosphate while a kidney cell differentiates to make *l*-amino acid oxidase. Each of these differentiations doubtless depends upon specific substances which activate or derepress the genes that make the differentiating proteins, for it now appears that many genes remain inactive unless specifically derepressed. This makes sense of the fact that when differentiated cells are removed from the body and cultured in isolation they commonly cease to make their characteristic proteins: removal from the body has removed them from the source of their genic derepressors. The technical achievements of Eagle, Puck, and others[11] in culturing mammalian cells *in vitro* will thus have to be matched by discovery of the relevant derepressors before this great new cellular technique can be fully exploited in the analysis of cell heredity and differentiation. However, the fact that the future can be defined in these terms only serves to emphasize the importance already attributable to variable genic activity in cell differentiation.

OTHER ASPECTS OF CELL DIFFERENTIATION

The simplest assumption to adopt concerning the basis of cell differentiation is that it is all due to one fundamental process. Since one process — variable genic activity and its regulation — is already known to be of wide applicability, this is clearly the candidate for the universal and exclusive basis of cell differentiation, if there is only one. To enable it to account not only for cell differences in the kinds and relative amounts of proteins — the most direct durable products of genic action — but also for cell differences in other substances and in the totality of cell structures and functions, one need only add the ancillary hypothesis of automatic self-assembly of the direct and indirect products of genic activity. According to this hypothesis, the gene-produced proteins interact by purely random collisions, as illustrated already in the formation of the tetramers of the lactic dehydrogenase isozymes. Just how far such automatic self-assembly can go in accounting for cell differentiation is of course not yet known. It is obviously good scientific procedure to refuse to multiply hypotheses until the facts demand it. Is there then any present compelling evidence that variable genic activity and self-assembly alone cannot account for some kinds of cell differences? If so, what other principle or principles are involved?

We may first of all eliminate certain superficial exceptions. For example, some cell differences are clearly due to the presence of viruses in the cells. But this is hardly a difference in principle, for these differentiations are doubtless traceable also to genes and their action, but the genes belong to the parasite or symbiont instead of to the host cell. In like manner, evidence is now accumulating that genic material, DNA, occurs in cytoplasmic structures such as plastids and perhaps mitochondria. Regardless of whether such DNA represents independent cytoplasmic genes or whether it is derived from nuclear genes, again no new principle beyond variable genic activity and self-assembly is yet obviously required.

Difficulties might be expected in proceeding from random collisions of molecules to non-random organization. But even here the recent elegant and penetrating studies from the laboratories of Kellenberger[12] and Edgar[13] on the control of virus organization indicate that genic action somehow also directs an amazing degree of precise non-random structural patterning. Yet a virus is far from a cell. Such analyses seem more relevant to the determination of cell organelles than to the determination of total cellular organization. A virus does not grow and divide like a cell. Its nucleic acid replicates and its other structures are separately formed, the parts later coming together in the final organization. On the contrary the integrity of non-random cell structure persists throughout growth and division which immediately suggests that the preexisting structure plays a decisive role that may not be explicable by mere random self-assembly of genic products.

Consider for example a highly organized asymmetric cell such as the protozoan, *Paramecium* (Figure 2). It divides transversely and yields two identically structured cells from one. This could not possibly be achieved by a mere transverse cut, for that divides it into two very different halves. The cell, however, achieves production of identical daughter cells by a complicated reproduction of all its precisely localized structures with distribution in such a way as to reconstitute the original pattern in both daughter cells.

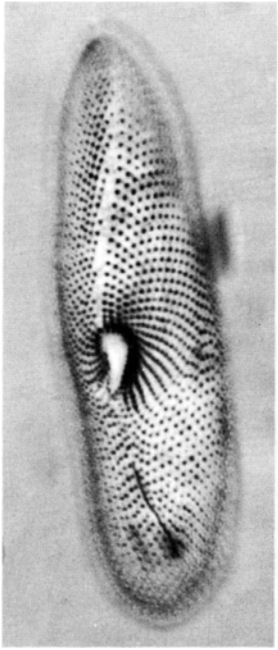
Although mere observation suggests that more than mere self-assembly of genic products is involved, experimental analysis is obviously required. What is needed is a thoroughly analyzed test case in which cells that are identical in the genes present and in the genes which are active differ not in the kinds of proteins and other substances and structures which are present, but only in their arrangement. Such differences have long been experimentally created and studied in some of the large, complex, unicellular, ciliated protozoa, especially in *Stentor* by Tartar[14] and in *Blepharisma* by Suzuki.[15] Comparable intracellular operations have recently been made on the egg of the amphibian, *Xenopus,* by Curtis.[16] All this work points in the same direction, but it lacks the final critical step of genetic analysis to test whether differences in genes or genic action were excluded. In fact that step has thus far been taken only in our work on *Paramecium*[17] (Figure 3).

We experimentally altered this precisely regular normal pattern of structure in various ways to yield, for example, cells with two or more cell mouths and gullets in various positions. You might suppose that such imposed upsets of normal cell structure would either be lethal or be rapidly corrected by the cell's genic actions, as the principle of self-assembly would lead one to expect. On the contrary, the bizarre cells are quite viable, the imposed differences persist, and they are as a rule inherited by the progeny at successive cell divisions and even through sexual reproduction. Moreover, what amounts to transplantation of nuclei in both directions between these and normal cells, as well as standard breeding analysis, showed that these hereditary cell differences are not due to differences in either the genes present or genic activity. They are due only to the differences in initial cell structure and organization.

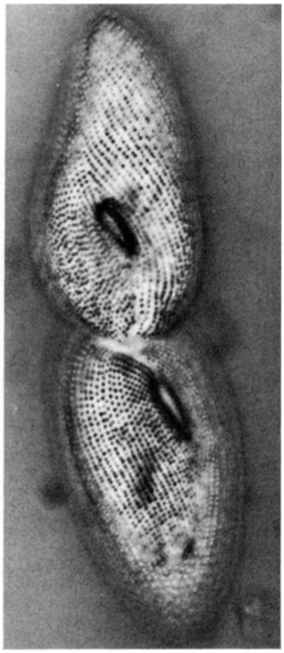
This is perhaps shown most simply and impressively by the simplest cell difference of all. Beisson[18] discovered how to create cells in which one or several of the 70-odd longitudinal rows of surface units was inverted and we have followed the fate of inversions. As Figure 4 shows, each unit is very asymmetrical. For example, the fiber emerging from the base of a cilium normally is on the animal's right and extends forward. The units of an inverted row have their fibers emerging on the left and extending backwards. Such experimentally produced changes have been perpetuated during fissions for over a year, during which more than 700 cell generations have taken place. As the length of a row is doubled in each cell generation, the original row (about 125μ long) has grown to $(2)^{700}$ or more than $(10)^{210}$ times its initial size. Had it been possible to keep all the progeny and place them end to end, the total length would now be roughly $(10)^{200}$ kilometers or about $(10)^{20}$ times the distance from the earth to the sun. This is perhaps enough to show the extreme stability and determinism of a merely structural intracellular rearrangement in the absence of differences in genes or gene action. It is a consequence of the fact that during growth new surface units appear within an existing row of units and are oriented in the same way as the units that are already there.[19] This ordering and arranging of new cell structure under the influence of preexisting cell structure I call "cytotaxis." Paul Weiss[20] has long adduced evidence for the same sort of thing, which he refers to as "macrocrystallinity."

This, I submit, is a second principle of cellular differentiation, one that is quite distinct from variable genic activity. The cell differences we have just been discussing are not characterized by different kinds of substances or structures, but by different numbers or arrangements of structures. Their perpetuation shows that they can be decisive in cell differentiation. Self-assembly of genic products alone cannot account for this. The place and orientation of the assembly of genic products are also determined by preexisting assemblies of molecules and structures. The picture that emerges is no less deterministic or molecular than self-assembly, but it is fuller and truer: there is more than "self" to the mechanism of assembly; it includes preexisting and independently modifiable assembly.

Thus, variable genic activity is decisive in cell differentiation by determining directly the kinds and proportions of molecular species present; but preexisting cellular structure is also decisive cytotactically by determining the location and orientation of these molecules and others formed from their reactions. As



a



b

FIGURE 2a. (*left*) Photo of pattern of rows of granules (silver impregnation method) marking the bases of the cilia in the unicellular animal *Paramecium aurelia*. The central comma-shaped space marks the position of the mouth and gullet; the long dark line in the lower part of the cell is the anus.

FIGURE 2b. (*right*) Same species, prepared in the same way, showing a stage of transverse division into two daughter cells. Note the re-creation of identical patterns and structures in both cells.

FIGURE 3. Photo of *Paramecium* with two mouths, two anuses, and two mid-ventral patterns of ciliary bases.

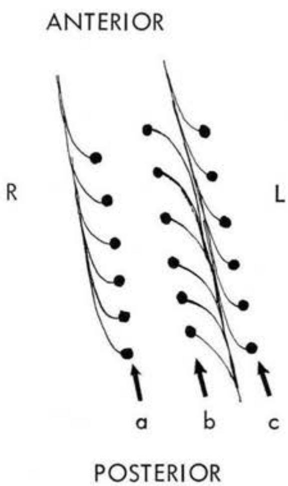


FIGURE 4. Diagram of small piece of the surface of *Paramecium* showing six surface units in each of three rows (*a*,*b*,*c*). In rows *a* and *c*, the fibers emerging from the ciliary base come off on the cell's right (*R*) and extend anteriorly (Ant.); in row *b*, the fibers come off on the left (*L*) and extend posteriorly (Post.). Normally, all rows are like *a* and *c;* but the abnormal orientation of a row (*b*), when present, is inherited.

Weiss[20] and Grobstein[21] have argued and as embryologists have long believed, the further one goes away from direct genic action in the economy of the cell and organism, the more important other factors become. While the genes determine the molecular building blocks and, through their properties, the kinds of molecular associations that can occur, the associations that actually do occur depend also on those that already exist. Calling these cytotactic events epigenetic or epigenic should not obscure their degree of independence or their decisiveness for the end result of cellular differentiation.

PROSPECTS

Where then are we heading in the further analysis of cellular differentiation?

The main gap in present knowledge about variable genic action is in the molecular species and events directly and indirectly involved in genic derepressions. It is especially important to discover the mechanisms of specificity by which one and not another gene is activated or repressed. After these things are known, it should become possible to understand the regular progressive *series* of genic activations and repressions which lead to normal developmental cellular differentiations. I have said nothing about interlocking pathways of metabolism with their feedback inhibitions, compensatory regulations, and stimulations, though they are obviously important in cellular function and differentiation. Much is already known about them. Much more needs to be discovered.

There are already strong indications that certain aspects of genic action in higher organisms are different from those in bacteria. This is not surprising in view of their more highly evolved chromosomes and nuclei. For example, genic messenger RNA is exceedingly short-lived in bacteria, of the order of minutes. Some results indicate much longer life of certain RNA messengers in vertebrates. The remarkable development of differentiations in the giant unicellular alga *Acetabularia,*[22] after removal of the nucleus, suggests that some of its messenger RNA may be stable for weeks.

Gibson and Beale's[23] remarkable results on an RNA intermediate between a gene and a trait in *Paramecium,* presumably a messenger RNA, seem to show that it is infectious and immortal. A mathematical analysis by Reeve and Ross[24] indicates that this RNA may have a weak capacity to multiply. Recently, Gibson and I[25] have introduced this RNA into another, not even closely related, cell (*Didin-*

ium). It then multiplies fast and without limit in the absence of the gene of *Paramecium* that apparently produced it.

Does this mean, in spite of the recent evidence for specific polymerases essential for the replication of each RNA virus, that messenger RNA's may also under certain conditions be capable of long persistence and even replicative reproduction? If so, this could be the basis of long persistent or permanent cellular differentiations in animals. On the other hand, there are also mechanisms in bacteria for rapid cellular destruction of messenger RNA and, in *Paramecium*, for nearly complete inhibition of multiplication of the one known to be capable of rapid multiplication in a foreign cell. Such inhibition would help to keep cells from becoming cluttered with the apparatus of previous differentiations. Higher plants have amazing capacities to develop whole plants from body cells or even from tumor cells, as Steward[26] and Armin Braun[27] have shown. Clearly organisms differ in the stabilities of their differentiations. These may be traceable at least in part to differences in the stabilities of their RNA messengers and/or their genic activations and repressions.

Perhaps the most important prospect for future theoretical work is to profit by the way in which success has been achieved in viral and bacterial work and to recognize that systems in microorganisms, in spite of their possible primitiveness, have much to tell about what goes on in higher organisms. The Monod-Jacob attack on the β-galactosidase locus in the colon bacterium, pursued for many years, gave us the concept of regulation of gene activity on which the major current attacks on cellular differentiation are based. Indeed, the successes and consequent possibility of applications of microbial work stem from concentrated attacks on relatively simple systems, often single genes. This should be a model for other work on cellular differentiation.

The slime mold mentioned earlier is such a simple system and concentrated work on it by K. Raper, John Bonner, Maurice Sussman, Barbara Wright, and others,[28] using genetic, chemical, and biological approaches has yielded very fruitful results. Recently some[29] have recognized the promise of analyzing the amazing synchronized and controllable differentiation of an amoeba into a flagellate, in protozoa such as a *Naegleria*.

In view of the central problem of genic control systems, I think it far better to select a simple cellular differentiation system in a cell that can be bred and subjected to standard breeding analysis. That is why we have used *Paramecium*.

Recently I[30] have found sexuality in, and have begun to exploit, a simpler system with the same advantages, a suctorian, *Tokophrya* (Figure 5). This cell under controllable conditions transforms in 3 minutes from a free-swimming ciliate into an immobile, attached, stalked tentacled cell. When the immobile complementary mating types are confronted at a distance each induces the other, apparently via diffusible substances, to assume amoeboid characteristics. I have

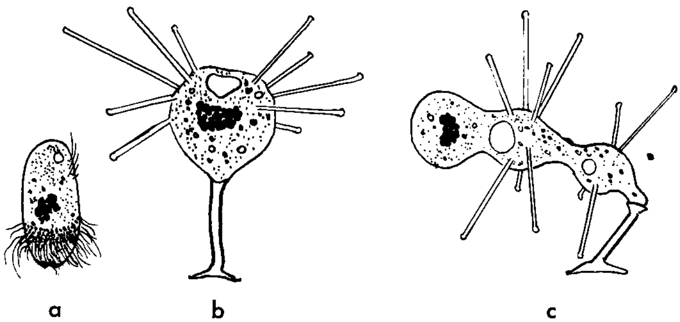


FIGURE 5. Diagram of three stages in life of the unicellular animal, *Tokophrya*. (*a*) Free swimming, ciliated "young." (*b*) Non-sexual phase of attached adult, lacking cilia, but bearing tentacles and stalk. (*c*) Sexual phase, with amoeboid features.

already a gene mutation affecting one differentiation and we — my student Laura Bukovsan and I — are discovering how to control other differentiations. This material seems extraordinarily favorable for deep penetration into problems of cellular differentiation.

With respect to future possible practical applications, knowledge of how to turn genes on and off and how to affect other cellular differentiations promises tremendous medical uses, as Tatum[31] long ago foresaw. It is not too much to imagine that tumor cells as well as other pathological forms of cells, perhaps even some aging changes, may prove reversible if we learn how to regulate the activities of the relevant genes. This basic level is obviously the one to attack, not the fully developed end results as non-scientists imagine, if we are to ameliorate the physical, and hence the total, well-being of the human individual. On both the theoretical and the applied sides, the analysis of cell differentiation — already brilliantly begun — indeed holds prospects full of promise for science and for man.

The Influence of The Environment

G. EVELYN HUTCHINSON

THE BIOSPHERE, or part of the earth within which organisms live, is a region in which temperatures range not far from those at which water is liquid. It receives a radiation flux of wavelength >3200 Å from the sun or has the products of photosynthesis made available by gravity as in the dark depths of the ocean. Numerous interfaces are present in most parts of the biosphere. It is geochemically characterized by atmophil elements in relative quantities such that both oxidized ($Eh \simeq 0.5$ volt) and quite reduced ($Eh \simeq 0.0$ volt) regions are both easily possible often within a few millimeters of each other as in lake sediments. Conditions for such a region on a planet are fairly critical, and would probably always involve loss of the initial gaseous phase with re-formation of the atmosphere from frozen or chemically combined material.

Chemically, living organisms are mainly made of cosmically common, light, easily soluble atmophil and lithophil elements. The special properties of the several important elements, such as hydrogen bridge formation, the formation of long carbon chains, the possible existence of $-COOH$ and of $-NH_2$ in the same molecule, the easily oxidized and reduced system $-SH\ HS- \rightleftharpoons -S-S-$, the high energy phosphate bond and various other less striking properties exhibited by common biophil elements (of which phosphorus, with an odd atomic number between Si and S, is the rarest) are obviously fundamental. Without these properties life as we know it would be impossible.

The reduction of magnesium concentration in the earth's crust, as compared with the mantle, and the consequent approximate equalization of the amounts of Na, K, Mg, and Ca provide a geochemical rather than a purely chemical example of the "fitness of the environment," to use Henderson's phrase. However, it is difficult to be sure we know we are talking sense in this field without comparative instances, or whether we are involved in problems like the insoluble metaphysical question of childhood, Why am I not someone else? The explora-

G. EVELYN HUTCHINSON Osborn Memorial Laboratory, Yale University

tion of the surface of Mars may give, long before the National Academy is celebrating another centenary, some welcome contrasting information. Meanwhile it is reasonable to suppose that extreme rarity and extreme insolubility, leading to a very low concentration of some elements within living tissues, do limit their functional importance.

If we consider an average mammalian liver cell of diameter about 23.4 μ, of volume, assuming a spherical form, of about 6700 μ^3, and of mass, if a density rather more than unity be assumed, of about 7×10^{-9} gm, we can obtain from published analyses[1] a rough idea of the mean number of atoms per cell as follows: —

$>10^{14}$	**H, O**	
10^{12}–10^{14}	**C, N**	
10^{10}–10^{12}	**S, P, Na, K, Mg, Cl, Ca, Fe,** Si	
10^{8}–10^{10}	**Zn,** Li, Rb, **Cu, Mn,** Al, Fe, Br	
10^{6}–10^{8}	Sn, Ti, Mo, **Co, I,** Pb, Ag, B, Sr, Ni, V, Sc, Cd, ?**Cr, Se**	
10^{4}–10^{6}	U, Hg, Be....	40 additional reactive natural elements
10^{2}–10^{4}	....	probably in these rows.
10^{0}–10^{2}	Ra	

The elements known to have a function in mammals, other than in maintaining the integrity of skeletal structures (as do F, and possibly[2] Sr and Ba), are given in bold-faced type. It is evident that the probability of an element having a function decreases with decreasing concentration. When such a table was published twenty years ago, there were 9 elements in the 10^6–10^8 atoms per cell category with only cobalt functional; now there are analytic data for 13 and a presumption that chromium and selenium, which with iodine are now known to be functional, fall here. Evidently about a quarter of the elements with 10^6–10^8 atoms per cell may have a function. In the next two rows we might guess probabilities of 0.1-0.01 of function. This may imply one or two surprises. What is interesting is that although cobalt is enriched relative to nickel in liver, as in nearly all tissues of higher animals, over its concentration in the lithosphere, or for that matter in plants, it is still little more abundant than lead and less so than molybdenum. To use an element such as cobalt, the biochemistry of utilization must be reasonably specific. There are plenty of atoms of various kinds around in such concentrations that they could play a part as antimetabolites as well as significant functional roles in enzyme systems. It is possible that this sets the lower limits of concentration at which biochemically significant substances occur. There might be too many commoner accidental and potentially interfering materials around for any very important substance to work practically at 10^4 atoms or molecules per cell. The variety of elementary composition may thus set the standard of purity within which biochemical evolution has occurred.

We may roughly divide most of the biosphere into a purely liquid part, and solid-liquid and solid-gaseous parts, corresponding to (1) the open ocean, (2) the sea bottom, margin, and inland waters, and (3) the colonized land surfaces. In the first, it is possible that iron, which is almost insoluble under oxidizing conditions in inorganic aqueous systems, usually limits the amount of living matter, while in the last, water supply is the most important determinant. In the water-solid systems, including lakes and neritic marine environments, phosphorus, nitrogen, and other elements may be limiting.

The whole plant community can, since it is interconnected through the CO_2 and O_2 of the atmosphere, be regarded as an extremely inefficient (efficiency not much more than 0.1 per cent in most cases in nature) photosynthetic machine. The details of the biochemistry of photosynthesis have been elucidated in recent years in most impressive studies by various investigators. We are, however, still rather ignorant of the quantitative details of the over-all biogeochemical process. It is evident that both the ocean[3] and the plant cover[4] of the continents play a major part in regulating the CO_2 content of the atmosphere, but this regulation is not sufficient to prevent a slow rise,[5] due partly, but perhaps not entirely, to the production of the gas by combustion of fossil fuels.

The use of plant material by animals as food is ordinarily a much more efficient process than the photosynthetic capture of the radiation flux from the sun, and allows the existence of a considerable mass of and extraordinary diversity of animals. This diversity is, however, clearly in part due to the diversity and structural complexity of higher plants. A large part of contemporary ecological research is devoted to elucidation of the general principles that permit the coexistence of very large numbers of species together in a single locality.

The complexity of communities has fascinated naturalists from before Darwin, who described it classically. Only recently has it become apparent what a wealth of quantitative relationships can be seen in the complex structure. Remarkable and quite diverse types of theory, some of which have proved of considerable value in empirical studies, have been developed to deal with this sort of problem, though we still have an enormous amount to learn. Ordinarily the principle of competitive exclusion,[6] which can be phrased in abstract geometrical terms as the statement that two coexisting species do not occupy the same niche, is a good point of departure so long as it is not applied in too naïve a manner. The claim may be made that the principle is inapplicable in practice because there would never be a possibility of demonstrating that the niche requirements of two species were, or were not, exactly the same. Actually what is involved is the question whether in competition two species could be so nearly equivalent that one would not replace the other in any reasonable time, such as the lifetime of an observer, the period of existence of relevant scientific records, or the period during which the average state of the habitat remained unchanged. The possibility that two competing species might be exactly matched contradicts what has been called the axiom of inequality,[6] that two natural bodies are never exactly the same. The possibility that in very large populations where the dynamics are essentially deterministic, they might be so nearly evenly matched that competition would proceed too slowly to detect, has been seriously suggested for phytoplankton associations, the multispecific nature of which seems otherwise paradoxical.[7] In this, as in other aspects of ecology, the role of time, or of the rates at which things can happen, has been inadequately studied.

In general the speed at which things happen in a very small organism depends on physical processes of which ordinary diffusion is likely to be the most critical. In larger organisms celestial mechanics of the solar system responsible for days, tidal periods, lunar months, and years, can introduce apparently arbitrary rates into the life of an organism which will interact with rates set by small-scale physical processes. The extent of various kinds of biological clocks is one of the most important phenomena recently discovered in biology. The evidence from odd cases in which periodic processes occur pathologically[8] suggests that the full significance of the gearing of organisms, including ourselves, to the cycle of day and night is not even yet apparent. Moreover a case perhaps could be made for supposing that sometimes clocks can exist which, having evolved in relation to the environment, are no longer set to be synchronous with external periodicities. The human female reproductive cycle, shared by a number of primates, has long suggested a lunar month in spite of the lack of synchrony. Possibly it represents a lunar clock no longer set by the moon; an effect of moonlight on the reproductive cycle of some tropical mammals, including prosimians, has been noted.[9]

Even more extraordinary are the clocks regulating the three species of seventeen-year and the three of thirteen-year cicadas.[10] Here apparently synchrony is adaptive, so that three species bear the brunt of predation, with the different broods emerging so irregularly that there is little chance of a permanent increase in predator population.

Individual cases of a striking kind can, as these, be given at least hypothetical, though very plausible, explanations. We still lack, however, a really clear understanding of the relationship of rates of living and of evolution to the rates of physical change in the universe.

Two extreme ways of evolution in relation to time are possible. Since natural selection will proceed faster when generations succeed each other faster, one way is the evolution of progressively smaller and more rapidly reproducing organisms. However, the smaller an organism the less it can do. An alternative path gives large, slower-reproducing organisms in which, when a nervous system capable of learning is developed, a premium is put on experience. Even in organisms such as plants, which do not learn in the ordinary sense of the word, a perennial can wait about at least metaphorically for a favorable season for reproduction. In the first case the time scale is set by the physical processes of diffusion; in the other extreme case, presumably by some function of the rate at which various things, such as learning, can occur.

In a varying environment, the time taken to learn about a seasonal or otherwise infrequent event will partly depend on the incidence of that event. Some sea birds seem to need several years' experience to learn how to get food for their chick or chicks.[11] The advantage of this learning must be great enough to offset the extra prereproductive mortality, which inevitably accompanies a delay in breeding.

It has recently been suggested that the great intraspecific competition on the limited feeding grounds near nesting sites puts a premium on expertness during the reproductive season and, until this has been acquired, attempts at breeding are wasteful and to some extent dangerous. Where mortality is lowest, possibly of the order of 3 per cent per annum in the albatrosses, the period prior to reproduction may be nine years, even though these immense birds have reached maximum size in their first year. In a case like this, learning can obviously only take place during the special period of reproduction, and expertness takes several years to acquire. But it is not clear why human learning of perennial activities should take so long when most of the individual events in our sensory and nervous systems take times measured in milliseconds, rather than months, while on the motor side, a good pianist can play ten notes a second if he has real cause to do so.

In the present and most legitimate excitement over the reading of genetic codes, it is important to remember that lexicography and grammar are not literature, even though the fixing of meaning to symbols and the rules of their ordering make literature possible. The literature of living organisms is very varied, and is perhaps most exciting in the epic or evolutionary forms in which organisms are continually changing in response to selection by a changing environment. Deduction from the possible molecular states of organisms is hardly likely to be an efficient way of exploration; an empirical approach to events is equally needed. A real ecology of time, relating the rates at which things happen in organisms, whether rapid physiological changes or the very slow changes of phylogenesis to the rates of the outside world, is so far only approached at the short-time physiological end. In the immediate future, as argon-potassium dating develops, it will be possible to study evolutionary rates of certain well known phyletic lines, notably in the Tertiary mammals, with greatly increased precision. Details of variation in evolutionary rates will become accessible and should add enormously to our knowledge of organic change under long time spans. This aspect of biology is likely to be one of immense importance as the Academy moves into its second century of high scientific endeavor.

The Evolution of Living Systems

ERNST MAYR

THE NUMBER, kind, and diversity of living systems is overwhelmingly great, and each system, in its particular way, is unique. In the short time available to me, it would be quite futile to try to describe the evolution of viruses and fungi, whales and sequoias, or elephants and hummingbirds. Perhaps we can arrive at valid generalizations by approaching the problem in a rather unorthodox way. Living systems evolve in order to meet the challenge of the environment. We can ask, therefore, what *are* the particular demands that organisms have to meet? The speakers preceding me have already focused attention on some of these demands.

The first challenge is to cope with a continuously changing and immensely diversified environment, the resources of which, however, are not inexhaustible. Mutation, the production of genetic variation, is the recognized means of coping with the diversity of the environment in space and time. Let us go back to the beginning of life. A primeval organism in need of a particular complex molecule in the primordial "soup" in which it lived, gained a special advantage by mutating in such a way that, after having exhausted this resource in its environment, it was able to synthesize the needed molecule from simpler molecules that were abundantly available. Simple organisms such as bacteria or viruses, with a new generation every 10 or 20 minutes and with enormous populations consisting of millions and billions of individuals, may well be able to adjust to the diversity and to the changes of the environment by mutation alone. In addition, they have numerous mechanisms of phenotypic adaptation. A capacity for mutation is perhaps the most important evolutionary characteristic of the simplest organisms.

More complex organisms, those with much longer generation times, much smaller population size, and particularly with a delicately balanced coadapted genotype, would find it hazardous to rely on mutation to cope with changes in the environment. The chances that the appropriate mutation would occur at the right time so that mutation alone could supply appropriate genetic variability for

ERNST MAYR Museum of Comparative Zoology at Harvard University

sudden changes in the environment of such organisms are virtually nil. What, then, is the prerequisite for the development of more complex living systems? It is the ability of different organisms to exchange "genetic information" with each other, the process the geneticist calls recombination, more popularly known as *sex*. The selective advantage of sex is so direct and so great that we can assume it arose at a very early stage in the history of life. Let us illustrate this advantage by a single example. A primitive organism able to synthesize amino acid *A*, but dependent on the primordial soup for amino acid *B*, and another organism able to synthesize amino acid *B*, but dependent on the primordial soup for amino acid *A*, by genetic recombination would be able to produce offspring with the ability to synthesize both amino acids and thus be able to live in an environment deficient in both of them. Genetic recombination can speed up evolutionary change enormously and assist in emancipation from the environment.

Numerous mechanisms evolved in due time to make recombination increasingly precise in every respect. The result was the evolution of elaborately constructed chromosomes; of diploidy through two homologous chromosome sets, one derived from the father, the other from the mother; of an elaborate process of meiosis during which homologous chromosomes exchange pieces so that the chromosomes of father and mother are transmitted to the grandchildren not intact, but as newly reconstituted chromosomes with a novel assortment of genes. These mechanisms regulate genetic recombination among individuals, by far the major source of genotypic variability in higher organisms.

The amount of genetic diversity within a single interbreeding population is regulated by a balance of mechanisms that favor inbreeding and such that favor outbreeding. The extremes, in this respect, are much greater among plants and lower animals than among higher animals. Extreme inbreeding (self-fertilization) and extreme outbreeding (regular hybridization with other species) are rare in higher animals. Outbreeders and inbreeders are drastically different living systems in which numerous adaptations are correlated in a harmonious manner.

The result of sexuality is that ever new combinations of genes can be tested by the environment in every generation. The enormous power of the process of genetic recombination by sexual reproduction becomes evident if we remember that in sexually reproducing species no two individuals are genetically identical. We must admit, sex is wonderful!

However, even sex has its drawbacks. To make this clear, let me set up for you the model of a universe consisting entirely of genetically different individuals that are *not* organized into species. Any individual may engage in genetic recombination with any other individual in this model. New gene complexes will be built up occasionally, as a result of chance, that have unique adaptive advantages. Yet, because in this particular evolutionary system there is no guarantee that such an exceptional individual will engage in genetic recombination *only* with individuals having a similarly adaptive genotype, it is inevitable that this exceptionally favorable genotype will eventually be destroyed by recombination during reproduction.

How can such a calamity be avoided? There are two possible means, and nature has adopted both. One method is to abandon sexual reproduction. Indeed we find all through the animal kingdom, and even more often among plants, a tendency to give up sexuality temporarily or permanently in order to give a successful genotype the opportunity to replicate itself unchanged, generation after generation, taking advantage of its unique superiority. The history of the organic world makes it clear, however, that such an evolutionary opportunist reaches the end of his rope sooner or later. Any sudden change of the environment will convert his genetic advantage into a handicap and, not having the ability to generate new genetic variability through recombination, he will inevitably become extinct.

The other solution is the "invention," if I may be pardoned for using this anthropomorphic term, of the biological species. The species is a protective system guaranteeing that only such individuals interbreed and exchange genes as have largely the same genotypes. In this system there is no danger that breakdown of genotypes will result from genetic recombination, because all the genes present in the gene pool of a species have been previously tested, through many generations, for their ability to recombine harmoniously. This does not preclude considerable variability within a species. Indeed, all our studies make us realize increasingly how vast is the genetic variability within even comparatively uniform species. Nevertheless, the basic developmental and homeostatic systems are the same, in principle, in all members of a species.

By simply explaining the biological meaning of species, I have deliberately avoided the tedious question of how to define a species. Let me add that the species can fulfill its function of protecting well integrated, harmonious genotypes only by having some mechanisms (called "isolating mechanisms") by which interbreeding with individuals of other species is prevented.

In our design of a perfect living system, we have now arrived at a system that can cope with the diversity of its environment and that has the means to protect its coadapted, harmonious genotype. As described, this well balanced system seems so conservative as to offer no opportunity for the origin of additional new systems. This conclusion, if true, would bring us into a real conflict with the evolutionary history of the world. The paleontologists tell us that the number of species has increased steadily during geological time and that the multiplication of species, in order to compensate for the extinction of species, must occur at a prodigious rate. If the species is as well balanced, well protected, and as delicate as we have described it, how can one species be divided into two? This serious problem stumped Darwin completely, and evolutionists have argued about it for more than one hundred years.

Eventually it was shown that there are two possible solutions, or perhaps I should say two normally occurring solutions. The first mode occurs very frequently in plants, but is rare in the animal kingdom. It consists in the doubling of the chromosome set so that the new individual is no longer a diploid with two sets of homologous chromosomes, but, let us say, a tetraploid with four sets of chromosomes, or if the process continues, a higher polyploid with an even higher chromosome number. The production of a polyploid constitutes instantaneous speciation; it produces in a single step an incompatibility between the parental and the daughter species.

The other mode of speciation is simplicity itself. Up to now, we have spoken of the species as something rigid, uniform, and monolithic. Actually, natural species, particularly those that are widespread, consist like the human species, of numerous local populations and races, all of them differing more or less from each other in their genetic composition. Some of these populations, particularly those at the periphery of the species range, are completely isolated from each other and from the main body of the species. Let us assume that one of these populations is prevented for a long time from exchanging genes with the rest of the species, because the isolating barrier — be it a mountain range, a desert, or a waterway — is impassable. Through the normal processes of mutation, recombination, and selection, the gene pool of the isolated population becomes more and more different from that of the rest of the species, finally reaching a level of distinctness that normally characterizes a different species. This process, called "geographic speciation," is by far the most widespread mode of speciation in the animal kingdom and quite likely the major pathway of speciation also in plants.

Before such an incipient species qualifies as a genuine new species, it must have acquired two properties during its genetic rebuilding. First, it must have acquired isolating mechanisms that prevent it from interbreeding with the parental species when the two again come into contact. Secondly, it must also have changed sufficiently in its demands on the environment, in its niche utilization (as the ecologist would say), so that it can live side by side with mother and sister species without succumbing to competition.

KINDS OF LIVING SYSTEMS

In our discussion of the evolution of living systems, I have concentrated up to now on major unit processes or phenomena, such as the role of mutation, of genetic recombination and sex, of the biological species, and of the process of speciation. These processes give us the mechanisms that make diversification of the living world possible, but they do not explain why there should be such an enormous variety of life on earth. There are surely more than three million species of animals and plants living on this earth, perhaps more than five million. What principle permits the coexistence of such a wealth of different kinds? This question troubled Darwin, and he found an answer for it that has stood the test of time. Two species in order to coexist must differ in their utilization of the resources of the environment in a way that reduces competition. During speciation there is a strong selective premium on becoming different from preexisting species by trying out new ecological niches. This experimentation in new adaptations and new specializations is the principal evolutionary significance of the process of speciation. Once in a long while one of these new species finds the door to a whole new adaptive kingdom. Such a species for instance was the original ancestor of the most successful of all groups of organisms, the insects, now counting more than a million species. The birds, the bony fishes, the flowering plants, and all other kinds of animals and plants, all originated ultimately from a single ancestral species. Once a species discovers an empty adaptive zone, it can speciate and radiate until this zone is filled by its descendants.

To avoid competition, organisms can diverge in numerous ways. Dr. Hutchinson has already mentioned size. Not only has there been a trend toward large size in evolution, but also other species and genera, often in the same lines, have evolved toward decreased size. Small size is by no means always a primitive trait.

Specialization for a very narrow niche is perhaps the most common evolutionary trend. This is the characteristic approach of the parasites. Literally thousands of parasites are restricted to a single host, indeed restricted to a small part of the body of the host. There are, for instance, three species of mites that live on different parts of the honey bee. Such extreme specialization is rare if not absent in the higher plants, but is characteristic for insects and explains their prodigious rate of speciation. The deep sea, lightless caves, and the interstices between sand grains along the seashore are habitats leading to specialization.

The counterpart of the specialist is the generalist. Individuals of such species have a broad tolerance to all sorts of variations of climate, habitat, and food. It seems difficult to become a successful generalist, but the very few species that can be thus classified are widespread and abundant. Man is the generalist par excellence with his ability to live in all latitudes and altitudes, in deserts and in forest, and to subsist on the pure meat diet of the Eskimos or on an almost pure vegetable diet. There are indications that generalists have unusually diversified gene pools and, as a result, produce rather high numbers of inferior genotypes by genetic recombination. Widespread and successful species of *Drosophila* seem to have more lethals than rare or restricted species. It is not certain that this observation can be applied to man, but this much is certain, that populations of man display much genetic variation. In man we do not have the sharply contrasting types ("morphs") that occur in many polymorphic populations of animals and plants. Instead we find rather complete intergradation of mental, artistic, manual, and physical capacities (and their absence). Yet, whether continuous or discontinuous, genetic variation has long been recognized as a useful device by which a species can broaden its tolerance and enlarge its niche. That the same is true for man is frequently forgotten. Our educators, for instance, have tended far too long to ignore man's genetic diversity and have tried to force identical educational schedules on highly diverse talents. Only within recent years have we begun to realize that equal opportunity calls for differences in education. Genetically different individuals do not have equal opportunities unless the environment is diversified.

Every increase in the diversity of the environment during the history of the world has resulted in a veritable burst of speciation. This is particularly easily demonstrated for changes in the biotic environment. The rise of the vertebrates was followed by a spectacular development of trematodes, cestodes, and other vertebrate parasites. The insects, whose history goes back to the Paleozoic nearly 400 million years ago, did not really become a great success until the flowering plants (angiosperms) evolved some 150 million years ago. These plants provided such an abundance of new adaptive zones and niches that the insects entered a truly explosive stage in their evolution. By now three-quarters of the known species of animals are insects, and their total number (including undiscovered species) is estimated to be as high as two or three million.

PARENTAL CARE

Let me discuss just one additional aspect of the diversity of living systems, care of the offspring. At one extreme we have the oysters that do nothing whatsoever for their offspring. They cast literally millions of eggs and male gametes into the sea, providing the opportunity for the eggs to be fertilized. Some of the fertilized eggs will settle in a favorable place and produce new oysters. The statistical probability that this will happen is small, owing to the adversity of the environment, and although a single full-grown oyster may produce more than 100 million eggs per breeding season, it will have on the average only one descendant. That numerous species of marine organisms practice this type of reproduction, many of them enormously abundant and many of them with an evolutionary history going back several hundred million years, indicates that this shotgun method of thrusting offspring into the world is surprisingly successful.

How different is reproduction in species with parental care! This always requires a drastic reduction in the number of offspring, and it usually means greatly enlarged yolk-rich eggs, it means the development of brood pouches, nests, or even internal placentae, and it often means the formation of a pair-bond to secure the participation of the male in the raising of the young. The ultimate development along this line of specialization is unquestionably man, with his enormous prolongation of childhood.

Behavioral characteristics are an important component of parental care, and our treatment of the evolution of living systems would be incomplete if we were to omit reference to behavior and to the central nervous system. The germ plasm of a fertilized egg contains in its DNA a coded genetic program that guides the development of the young organism and its reactions to the environment. However, there are drastic differences among species concerning the precision of the inherited information and the extent to which the individual can benefit from experience. The young in some species appear to be born with a genetic program containing an almost complete set of ready-made, predictable responses to the stimuli of the environment. We say of such an organism that its behavior is unlearned, innate, instinctive, that its behavior program is closed. The other extreme is provided by organisms that have a great capacity to benefit from experience, to learn how to react to the environment, to continue adding "information" to their behavior program, which consequently is an open program.

Let us look a little more closely at open and closed programs and their evolutionary potential. We are all familiar with the famous story of imprinting explored by Konrad Lorenz. Young geese or ducklings just hatched from the egg will adopt as parent any moving object (but preferably one making appropriate noises). If hatched in an incubator, they will follow their human caretaker and not only consider him their parent but consider themselves as belonging to the human species. For instance, upon reaching sexual maturity they may tend to display to and court a human individual rather than another goose. The reason for this seemingly absurd behavior is that the hatching gosling does not have an inborn knowledge of the Gestalt of its parent; all it has is readiness to fill in this Gestalt into its program. Its genetically coded program is open; it provides for a readiness to adopt as parent the first moving object seen after hatching. In nature, of course, this is invariably the parent.

Let us contrast this open program with the completely closed one of another bird, the parasitic cowbird. The mother cowbird, like the European cuckoo, lays her eggs in the nests of various kinds of songbirds such as yellow warblers, vireos, or song sparrows, then abandons them completely. The young cowbird is raised by his foster parents, and yet, as soon as he is fledged, he seeks other young cowbirds and gathers into large flocks with them. For the rest of his life, he associates with members of his own species. The Gestalt of his own species is firmly imbedded in the genetic program with which the cowbird is endowed from the very beginning. It is — at least in respect to species recognition — a completely closed program. In other respects, much of the behavioral program of the cowbird is open, that is, ready to incorporate experiences by learning. Indeed, there is probably no species of animals, not even among the protozoans, that does not at least to some extent derive benefit from learning processes. On the whole, and certainly among the higher vertebrates, there has been a tendency to replace rigidly closed programs by open ones or, as the student of animal behavior would say, to replace rigidly instinctive behavior by learned behavior. This change is not a change in an isolated character. It is part of a whole chain reaction of biological changes. Since man is the culmination of this particular evolutionary trend, we naturally have a special interest in this trend. Capacity for learning can best be utilized if the young is associated with someone from whom to learn, most conveniently his parents. Consequently there is strong selection pressure in favor of extending the period of childhood. And since parents can take care of only a limited number of young, there is selection in favor of reducing the number of offspring. We have here the paradoxical situation that parents with a smaller number of young may nevertheless have a greater number of grandchildren, because mortality among well-cared-for and well-prepared young may be reduced even more drastically than the birth rate.

The sequence of events I have just outlined describes one of the dominating evolutionary trends in the primates, a trend that reaches its extreme in man. A broad capacity for learning is an indispensable prerequisite for the development of culture, of ethics, of religion. But the oyster proves that there are avenues to biological success other than parental care and the ability to learn.

One final point: how can we explain the harmony of living systems? Attributes of an organism are not independent variables but interdependent components of a single system. Large brain size, the ability to learn, long childhood, and many other attributes of man, all belong together; they are parts of a single harmoniously functioning system. And so it is with all animals and plants. The modern population geneticist stresses the same point. The genes of a gene pool have been brought together for harmonious cooperation, they are coadapted. This harmony and perfection of nature (to which the Greeks referred in the word *Cosmos*) has impressed philosophers from the very beginning. Yet there seems to be an unresolved conflict between this harmony of nature and the apparent randomness of evolutionary processes, beginning with mutation and comprising also much of reproduction and mortality. Opponents of the Darwinian theory of evolution have claimed that the conflict between the harmony of nature and the apparent haphazardness of evolutionary processes could *not* be resolved.

The evolutionist, however, points out that this objection is valid only if evolution is a one-step process. In reality, every evolutionary change involves two steps. The first is the production of new genetic diversity through mutation, recombination, and related processes. On this level randomness is indeed predominant. The second step, however — selection of those individuals that are to make up the breeding population of the next generation — is largely determined by genetically controlled adaptive properties. This is what natural selection means; only that which maintains or increases the harmony of the system will be selected.

The concept of natural selection, the heart of the evolutionary theory, is still widely misunderstood. Natural selection says no more and no less than that certain genotypes have a greater than average statistical chance to survive and reproduce under given conditions. Two aspects of this concept need emphasis. The first is that selection is not a theory but a straightforward fact. Thousands of experiments have proved that the probability that an individual will survive and reproduce is not a matter of accident, but a consequence of its genetic endowment. The second point is that selective superiority gives only a statistical advantage. It increases the probability of survival and reproduction, other things being equal.

Natural selection is measured in terms of the contribution a genotype makes to the genetic composition of the next generation. Reproductive success of a wild organism is controlled by the sum of the adaptive properties possessed by the individual, including his resistance to weather, his ability to escape enemies, and to find food. General superiority in these and other properties permits an individual to reach the age of reproduction.

In civilized man these two components of selective value, adaptive superiority and reproductive success, no longer coincide. The individuals with above average genetic endowment do not necessarily make an above average contribution to the gene pool of the next generation. Indeed the shiftless, improvident individual who has a child every year is sure to add more genes to the gene pool of the next generation than those who carefully plan the size of their families. Natural selection has no answer to this predicament. The separation in the modern human society of mere reproductive success from genuine adaptedness poses an extremely serious problem for man's future.

In this brief discussion of the evolution of living systems, I have been unable to do more than outline basic problems. We are beginning to understand the role of mutation, of genetic recombination, and of natural selection. The comparative study of the overwhelming multitude of diverse living systems has only begun. Because much of our environment consists of living systems, their study is of great importance. Indeed it is a prerequisite for understanding ourselves, since man also is a living system.

*

Physiological and Cultural Determinants of Behavior

NEAL E. MILLER

THE PRECEDING lectures have presented an exhilarating panorama of man's scientific discoveries ranging from the creation of the elements during cataclysmically violent explosions in the ancient vast galaxies of space to the more recent evolution of living systems on our own tiny temperate earth. I have been given the task of making a transition from these topics to those for the next session. How is it that the organisms developing from certain complex helical molecules of DNA will be standing on this platform to talk to you about problems involved in the communication of scientific knowledge, its relationship to public policy, and the satisfaction of human aspirations?

I cannot answer this question, but I can sketch for you certain relevant steps toward an answer. These will also serve to illustrate a few of the many diverse recent advances in the broad area of behavioral science.

Behavioral scientists are beginning to learn something of how the brain works. The human brain has at least 10^{10} neurons, many of which have thousands of complex branches, resulting in a total of at least 10^{13} (a million times 10 million) synaptic connections. Many of these myriad connections appear to be innately determined. They provide an essential basis for aspects of everyday behavior which all of us casually take for granted. One of these is recognizing a horizontal line. This is not as simple as one might think, because as one's head and eyes move, the same horizontal line stimulates quite different arrays of points on the retina. We are beginning to learn something of how such perception is achieved.

Using microelectrodes to record the activity of individual neurons, Hubel and Wiesel[34] find in the visual cortex single nerve cells that fire whenever a specific horizontal line on the retina is illuminated. Presumably the many different receptors stimulated by the points on this specific line are each connected to this cell.

NEAL E. MILLER Department of Psychology, Yale University

Other cells fire when specific horizontal lines in other locations are illuminated. These cells do not respond to general changes in illumination or to small patches on the line. Yet other cells will fire whenever *any* horizontal line is illuminated. Presumably each of these neurons receives connections from many cells each responding to different horizontal lines. Because of the uniformity with which such cells are arranged in orderly columns, their connections presumably are largely innately pre-determined, although experience is necessary for certain other aspects of complex visual perception.[24]

Similar arrangements are found for responding to lines oriented in other directions and to a considerable number of other salient features of the visual world. This work begins to give us an idea of how a multitude of specific connections in the brain abstract information from the stimulation of individual rods and cones and allow us to respond to more general attributes, such as a line running in a specific direction.

You have seen how one aspect of the innate organization of the brain provides a basis for everyday behavior. There are many other such examples. Even at the level of human social behavior, we may have many organized patterns, analogous to those which have been studied by the ethologists, which have not yet been recognized as instinctual because they are overlaid by social learning. I agree with previous speakers that innate behavior should continue to be vigorously investigated.

Our understanding of the mechanism of innate organization of the nervous system is being increased by recent research on lower animals. In one type of work Sperry and his associates cut the optic nerve, manipulate it in various unusual ways, and then observe the process of regeneration.[4, 5] When the eye of a frog is turned upside down, the fibers regenerating from the nerve cell in the retina do not grow along the nearest old path or remain in the same spatial relationship to each other. They are not guided by adaptive function. Instead, each of the 400,000 nerve fibers curves around to establish essentially the same connection that it had before, producing an orderly visual field which in this case is maladaptively inverted so that for the rest of its life the frog's tongue will snap downward when a fly buzzes by above. Recent experiments on the regeneration of cut optic nerves of fishes have shown that, even when the fibers are forced to detour and pass through abnormal routes in the brain, they will curve back to arrive at their preordained destination.

Such results make a number of older theories untenable and show that various parts of the embryo must differentiate into some fantastically specific coding system with appropriate matches between the more than a million receptors in the peripheral sense organs and the corresponding cells in the brain so that each of the connecting fibers will be guided to grow to the correct destination. Reflection on how a single fertilized egg cell can develop to produce these results gives one a deeper appreciation of the wonderful process of differentiation and orderly growth which has been discussed in Professor Sonneborn's illuminating lecture.[66]

You have had a glimpse of one kind of research that is helping us to understand how the brain is organized and functions in the processing of information. Research on this and other functions is giving us a new picture of the brain. We know that the all-or-none conduction of the long nerve fibers called axones does not apply to the synapses which connect one nerve cell with another and are especially prevalent in the brain. Here the transmitting tip of an axone acts like a gland to secrete a chemical transmitting substance, such as acetylcholine, to achieve a finely graded response which may be either excitatory or inhibitory. These conclusions are founded on converging evidence from many techniques. Some of this work has been done by biophysicists who thrust micropipettes with several barrels into a single nerve cell, using a conductive solution in one pipette to record the electrical activity of the cell, while minute quantities of various chemicals are injected electrophoretically via the other barrels. Studies with the electron microscope have verified other details. Yet other studies have used a push-pull cannula to wash out and measure for a group of nerve cells the greater production of the transmitter, acetylcholine, when they are active than when they are not.[18, 22]

The behavioral effects of stimulating considerable populations of cells in certain locations of the brain have been studied by Grossman in my laboratory.[25] He implanted a tiny cannula under anesthesia through which chemicals can later be introduced to specific sites in the brain of the normally behaving, unanesthetized animal. His studies have shown that after a rat has been thoroughly satiated on both food and water, injecting a minute amount of acetylcholine or of carbachol directly into a certain part of the brain will cause it to drink, while epinephrine or norepinephrine injected into the same site will cause the same satiated rat to eat. A series of control studies support the most obvious interpretation of these results, namely, that the neuromechanisms involved in the motivations of hunger and thirst are chemically coded.[26, 52]

Other fundamental changes in our ideas of the brain come from a study of the reticular formation, the central core of the brain which is characterized by a multitude of short interconnections in contrast with the longer fibers involved in the classical, more peripheral, sensory and motor systems which have long been understood. The study of this system and related nuclei has changed our concept of sleep. We now know that there is a wakefulness region of the brain. If this is damaged, the animal remains permanently somnolent. But there are also two different sleep centers. Instead of activating the animal, stimulation of these centers puts him to sleep. One center is responsible for a distinctive form of light sleep and the other for deep sleep. We no longer think of sleep as an over-all decrease in the activity of the brain. We know that the rate of spontaneous firing of many cells is increased during sleep. The decreased responsiveness to external stimulation seems to be a change in the signal-to-noise ratio, which is produced more by increased noise than by decreased signal.[43]

The brain used to be thought of as a passive switchboard, activated only when stimulated by the peripheral sense organs. Now we know that each signal impinges on a background of ongoing activity, is modified by this activity, and in turn modifies it. In addition to the pathways coming in from the sense organs are fibers carrying impulses from the brain to the sense organs. These impulses from the brain can either increase the rate of firing of a specific class of sense organs or decrease it. Impulses from the brain can also affect various relay points between the sense organs and the highest levels of the brain. In this way the brain can control its own input.[41]

The mechanism I have just described gives us a physiological basis for some of the psychological phenomena of attention, an area in which there has been a great deal of recent research at the purely behavioral level. This purely behavioral research has discovered many lawful relationships for which the brain mechanisms have not yet been identified, although carefully controlled experiments have demonstrated that many of the phenomena, such as your capacity to listen selectively to either one of two equally loud conversations at a cocktail party, must occur centrally in the brain rather than peripherally in the sense organs.[9]

In addition to its obvious relationship to the eyes, ears, nose, and taste buds that sense the external world, the brain is now known to contain within itself specialized receptors for sensing the internal state of the body. For example, by using a thermode to heat or cool a tiny specific region of the anterior hypothalamus, which is a primitive part of the brain, and by recording from there with microelectrodes, Nakayama, Hammel, Hardy, and Eisenman[57] have found that the majority of neurons are relatively unaffected by moderate changes in their temperature. However, there are some neurons here that increase their rate of firing when they are slightly heated, and others that increase their rate when they are slightly cooled. These cells seem to serve as, or be connected to, specialized "sense organs" for measuring small changes in the temperature of the surrounding blood.

Heating this region of an animal's brain causes panting and increased blood supply to the skin which serve to lower the body temperature. Cooling it causes the opposite effect of shivering and decreased blood supply to the skin.[35] It also stimulates the secretion of the thyroid, which in turn speeds up the body's burning of fuel.[2] In the experiments in which only this tiny region of the brain is cooled, these effects produce a fever, but when the whole body is cooled under normal conditions, they serve to restore the animal's temperature to normal. This is one example of a homeostatic mechanism that causes the internal environment to be held at an optimal level, in this case a constant level of temperature.

The regulative effects that I have just described are relatively direct. But the beautiful picture is extended still further by the fact that cooling this region of the brain will make a satiated animal hungry, so that it will eat, whereas heating this region elicits drinking.[3] Thus, this temperature-regulating mechanism is tied in with hunger and thirst which motivate behavior that helps the animal to anticipate its needs for the fuel it will burn to keep warm, or the water it will evaporate to cool off.

In short, a whole series of homeostatic mechanisms ranging from changes in metabolism to the motivation of the behavior of seeking food or water is touched off by the cells in the brain that respond to temperature.

A considerable number of other homeostatic mechanisms are known to be involved in the motivation of specific types of behavior. For example, certain receptors in the brain respond to osmotic pressure so that a minute injection into the proper place in the brain of a solution that is slightly more salty than body fluid will motivate animals that have just been satiated on water to drink and also to perform responses that they have learned to get water. Under normal circumstances such drinking reduces the salinity of the body fluids down to its proper level. Conversely, a minute injection of water will cause a dehydrated animal to stop drinking or working for water.[46, 58]

Yet other cells of the brain respond to specific hormones so that activities such as nest-building in rats can be elicited by injecting a minute quantity of the proper hormone into the correct site in the brain.[20] Hormones can also exert an inhibitory effect on specific areas of the brain without affecting other areas. Thus the progestational steroids normally secreted by the ovary during pregnancy inhibit that part of the hypothalamus responsible for stimulating the pituitary to release the gonadotropic hormone which leads to ovulation. This hormone does not inhibit, however, other parts of the brain which are involved in sexual motivation, sexual performance, or sexual pleasure. A similar synthetic compound is used in oral contraception, a technique which would be considerably less attractive if this hormone's action on different parts of the brain were less specific.[63]

Motivation has been mentioned a number of times. One of the exciting developments in recent years has been the rapid advance in our understanding of the brain mechanisms involved in certain basic motivations such as hunger and thirst. A combination of physiological and psychological techniques has been especially fruitful. In Figure 1 you will see a rat that has learned when hungry the habit of pushing back a little panel to get food. Since this rat has just been thoroughly satiated on food, it is not performing this habit. But as soon as the experimenter presses a key to stimulate a feeding center in the rat's brain with approximately 15-millionths of an ampere of current, hunger is elicited so that the rat promptly performs the learned habit of going directly to the right place and pushing open the panel to get the pellets of food hidden behind it. He does this repeatedly. As soon as the brain stimulation is turned off, the rat stops performing this habit. Stimulation of this part of the brain can also motivate a rat which has just been satiated on food to learn where to find food. Different tests of this kind prove that such stimulation elicits more than reflex eating; it functions just like hunger in motivating different types of learned food-seeking behavior.[48]

Another advance has been the development of purely behavioral techniques for the laboratory study of non-homeostatic motivations, such as curiosity, the affection of the infant monkey for its mother, and the desire of a human child for social approval.[6, 23, 29] Progress in such research may be expected to continue; eventually a great deal more will be known about how to stimulate, instead of stultify, intellectual curiosity in the classroom and how to arrange the social situation so that the strong influence of the classmates will favor, instead of conflict with, the goals of the parents and teachers.[10, 15]

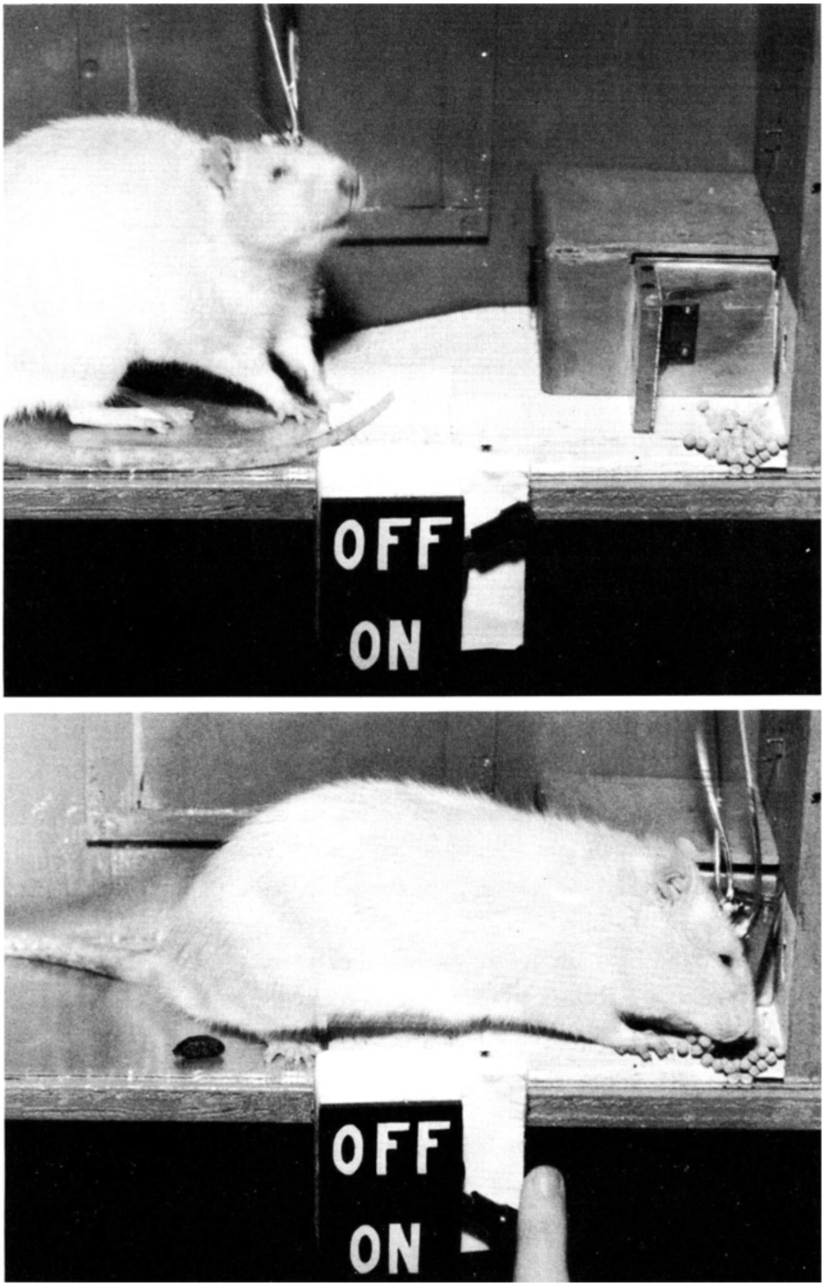


FIGURE 1. Electrical stimulation in the "feeding area" of the lateral hypothalamus causes a rat, which was satiated and hence previously inactive, to go promptly to the right place and perform the learned habit of pushing back a hinged panel to get the pellets of food hidden behind it. (Photographed in author's laboratory.)

I have sampled our scientific understanding of how the brain processes information, how it exerts a control over its own input, how it contains sense organs to help it to regulate the internal environment, and how it is involved in motivation. Let me turn now to certain recent knowledge about how we react to stress. Under stress, certain parts of the hypothalamus function as a gland to secrete substances which are carried by the blood stream and excite specific sites on the pituitary gland, causing it to secrete ACTH, which in turn causes the cortex of the adrenals to excrete ACH, which is one of the factors that causes the stomach to secrete hydrochloric acid. Under extreme circumstances, this acid and other effects of the complex reaction to stress are involved in producing ulcers in the stomach. Experimental studies have shown that direct electrical stimulation of appropriate areas of the hypothalamus can produce such ulcers.[21]

Clinical observations suggest that chronic psychological stress can produce ulcers. These observations are confirmed by a variety of experiments in which various animals have been subjected to situations involving difficult decisions, or conflict between approaching a place to get food when hungry and staying away from it to avoid electric shocks. These experiments have shown that such psychological conditions, which presumably produce stress by activating the hypothalamus via the cortex, can elicit the pituitary-adrenal mechanism just described, increase the secretion of hydrochloric acid, and also produce ulcers.[49, 59]

The same general stress mechanism seems to be involved also, along with many other factors such as the level of cholesterol and sodium chloride, in the development of arteriosclerosis, and in the production of high blood pressure. Such conditions have been produced by prolonged electrical stimulation of the hypothalamus and also by subjecting animals to psychological stresses which presumably initiate a similar chain of physiological reactions.[27, 53] Other studies are indicating that cultural conditions producing psychological stress can contribute to the physical symptoms of chronic high blood pressure. Scotch[64] has shown that members of the Zulu tribe, living in villages in which their culture is relatively unaltered, have a very low rate of hypertension. Those who have been moved to cities where their old culture patterns have been disrupted have an excessively high rate. For both groups, the occurrence of hypertension is correlated with sociological variables presumably involved in stress, but these are different under the two sets of cultural conditions. Somewhat similar correlations between sociological conditions producing stress and hypertension have been suggested by studies of the Negroes in the Chicago area.[69] These are pioneering studies; their results must be accepted with some caution. More penetrating studies of the roles of sociological variables in psychosomatic symptoms are in progress.

As more work is being done on various aspects of stress, the convergence among the results of experimental studies by physiologists, clinical observations by psychiatrists, experimental studies by psychologists, and cultural studies by sociologists and anthropologists is becoming increasingly impressive. The results from different disciplines are beginning to fit together into a significant pattern. Additional parts of the puzzle lie within our reach, but we have just begun to grasp them.

Thus far, we have primarily sampled recent knowledge about innate physiological determinants of behavior. As Professors Hutchinson and Mayr have told you, one of the important innate capacities of higher animals is the ability to learn from experience.[36, 45] Recent advances in our knowledge of learning, memory, and reasoning merit an entire separate lecture, but there is not time.[11, 14, 16, 19, 32, 37, 42, 47, 50, 51]

The great development of the human brain, which is disproportionately large in the areas representing the vocal apparatus, has been associated with a unique ability to learn and use language. Through language, the experience of one generation is passed on as a cultural heritage to the next. For long-lived organisms, the process of natural selection is extraordinarily slow. Within the individual's lifetime, the process of learning directly from the environment is much faster, but still limited. When the cumulative effects of the lifetime learning of multitudes of different individuals in each generation can be passed on as part of the culture, however, developments proceed with an entirely different order of magnitude of ever accelerating speed.

Man's mode of adjustment to the environment is preeminently cultural. When a lower animal moves into a colder ecological niche, it slowly evolves the capacity to grow a thicker fur; when man moves into a much colder climate, he invents warmer clothes and a better house.

For the ancient men who had a much poorer cultural heritage, the process of innovation was vastly slower. Man has been in existence for approximately a million years. It is unlikely that his brain has evolved significantly in the last 50,000 years, and perhaps not for much longer. But man did not develop agriculture until 10,000 years ago. It took him about 5,000 more years to invent the wheel as a means of transportation, which as far as we know was invented only once and was never discovered by the Indians of North and South America, in spite of the fact that the Incas had paved roads![13] The steam engine as a source of industrial power was invented two hundred years ago, the aeroplane within the lifetime of many of us in this room, and atomic energy and space vehicles still more recently. Compare the glacially slow initial progress with the torrent which you have seen during your own life. This is the effect of the cumulative heritage of culture.

Science is a recent product of this cultural development, one which is powerfully accelerating the process. As President Seitz pointed out during the second session, science flourishes only under certain cultural conditions. Remember what happened to Galileo for advocating the heliocentric hypothesis, how the early anatomists were prevented from dissecting cadavers, and the present activities of antivivisectionists. Modern science would be impossible without an advanced educational system and a strong economic base. Without an enormous cultural heritage, members of the human species would never be giving the papers that are included in the next session on The Scientific Endeavor.

To understand man we must study his culture as well as his physiology and psychology. The basic *principles* of human learning and behavior depend on the innate physiological structure of man. But in many cases we have not yet been able to relate these principles to the innate physiology and have found it most useful to study them at the purely psychological, or in other words, behavioral level, and the discovery of the links to physiology will not eliminate psychology any more than the quantum mechanics of valences has eliminated the discipline of chemistry. On the other hand, the crucial *conditions* of human learning and behavior are being scientifically studied by anthropologists, sociologists, and other social scientists. In order to understand human behavior, one must know both the principles and the conditions.[17] That is why the scientific study of culture is important.

It is often said that human behavior is unpredictable. This is not true. Under appropriate cultural conditions there is a high degree of predictability. Without such predictability, civilization would be impossible. Look around you. It is safe to predict that no one will be sitting there naked. You might stop for a moment to think how much someone would have to pay you in order to make you undress at this meeting. This will give you some idea of the power of culture.

How bad a pest would some elderly relative of yours have to be before you would kill him? That is the strength of the cultural taboo in our society against murder. In India, cows, monkeys, parrots, insects, and other pests ravage the crops of an undernourished population making it extremely difficult to introduce efficient methods of modern agriculture which involve growing large fields of the same crop and hence invite multiplication of the pests that prey on that crop. Since the Hindu believes that any one of these animals might be the reincarnation of one of his relatives, he will not kill any of these pests. For him the sacredness of all animal life is more important than his own hunger for food. On the other hand, the Eskimo used to abandon his parents to certain death when they were too old to care for themselves, and the aged parents asked to be killed in this way.[7] These are but a few examples of the enormous power of the social conditions of culture.

The first emphasis of the social anthropologists was on the extreme diversity of cultures. They proved that an astonishing amount of human nature is the product of cultural conditions and hence is not the same everywhere. For example, there are some cultures in which different tribes have interacted for long periods without warfare, and in some cases even without military weapons. Peace is compatible with human nature![31, 54, 67, 71]

Even within our own American society, we have various subcultures. In the different social classes, the ideas of what is right and what is wrong, the values placed on education, having an illegitimate child, aggression, saving money, and many other aspects of life are far more different than most of us realize.[70] There also are regional differences in culture, as illustrated by recent events in the deep South. Certain occupational groupings are minor subcultures. Everyone takes his own immediate culture for granted, and greatly underestimates the degree to which others react differently, because everyone tends to associate with people from a similar cultural background.

As they have advanced beyond their initial descriptions, anthropologists have found that the fact that many radically different ways of doing things in other cultures are quite possible does not mean that it is a matter of indifference how things are done in a given culture. They have learned much about how the different parts of a given culture dovetail in a lawful, functional way, so that changes of certain aspects may be strongly resisted, and if forced, may produce repercussions throughout the society. To give a simple example, in the Trobriand Islands, instead of a husband supporting his wife, a brother supports his sister, and if his sister has the status of being married to a chief, she must be well provided for. Each chief had many wives which were the source of his economic power, enabling him to supply food and drink for work parties which accomplished essential civil projects such as hollowing out great logs to produce ocean-going canoes. When pressure from colonial administrators caused abandonment of polygamy, the chiefs lost the extra income from having several brothers-in-law, they were too poor to throw work parties, their prestige and authority declined, and necessary civic tasks were neglected. A change in one aspect of the society had unexpected, far-reaching effects.[44]

An additional example of the functional interdependence of various aspects of a culture comes from Murdock's[55] comparative study of 250 societies which also illustrates the modern trend toward quantitative work. In a statistical analysis he has found that different patterns of sexual taboos in these societies are highly correlated with the ways in which kinship terms, such as "Mother" or "Aunt," are applied to larger or smaller classes of maternal or paternal relatives. The various uses of kinship terms are related to family living arrangements — whether the bride moves in with the husband's relatives, the husband moves in with the bride's relatives, or both go off independently. These customs are in turn dependent upon economic behavior, for example, hunting *versus* agriculture. Murdock also has shown how those changes which are known to have occurred during the history of these societies progressed in a lawful sequence with the changes in economic behavior coming first, followed by changes in living arrangements, and then in kinship systems.

We have seen that cultural changes are occurring at an accelerated rate. There is less time now for different aspects of a society to adjust by the slow, blind processes of social trial-and-error which were sufficient in the past. Some of the later speakers may be dealing with problems of this kind in discussing new relationships between science, government, and the universities.

In the studies of the hypertension of the Zulus, we saw one symptom of the stress involved in changing from one culture pattern to another. Other effects of the disorganization produced by relatively milder changes are illustrated by the fact that in our own country problems of delinquency and crime have characteristically involved the children of the newest wave of immigrants. Yet other studies seem to show that certain sociological variables, especially social disorganization, are related to a higher rate of mental illness.[33, 39, 56, 62, 68]

The functional relationships among the different parts of a society become especially important when attempts are made to introduce rapid technological changes in so-called "underdeveloped" countries. The changes occurring in such countries today create an urgent need for us to apply what we know, and at the same time present unparalleled opportunities for increasing our scientific understanding of the dynamics of rapid social change.

In recent years some political scientists and economists have greatly broadened their perspective by studying other societies and cultures, including some of the technologically underdeveloped ones. Instead of concentrating almost exclusively on official documents or government statistics, they have devised techniques for studying political and economic behavior in the field. Unfortunately, these developments, which are changing these subjects from what they were when most of us studied them in college, are beyond the scope of this paper.[1, 12, 28, 60, 61]

I shall conclude with one more example illustrating the way in which different disciplines, ranging from the physiological to the cultural, have profitably converged on a behavioral problem. You will remember that culture must be learned by children and that such learning during childhood is a crucial link in passing it on from generation to generation. In addition to this general consideration, clinical observations have suggested that there may be certain critical periods in childhood during which experiences may have an especially profound and long-lasting effect.

These observations have been supported by recent experimental studies on animals. The imprinting, which Professor Mayr[45] has described, can occur only during a limited period of the newly hatched bird's life. Birds will not learn to follow an object if their first exposure is after the critical period. Female rats restrained from grooming themselves during a certain period of infancy do not show normal nest building in later life and will eat, instead of care for, their young. Similarly, monkeys reared with artificial mothers do not display normal maternal behavior in later life, and if isolated from other infant monkeys during a certain period of childhood will not acquire normal sexual behavior when adult. They also show other strikingly neurotic symptoms. Bottle-fed sheep do not develop gregariousness, but graze alone. That genetic factors are also involved, however, is shown by the fact that not all mammals become as gregarious as sheep.[65]

The critical periods involve the interaction of developing innate patterns with learning. Their occurrence is firmly established in certain lower animals but we need to know much more about the detailed processes involved, and their occurrence in human infants.

Evidence that there are such critical periods in the development of the human infant has been secured by taking advantage of experiments of nature in which illness or other unusual circumstances have forced the child to be separated from its mother and mother-surrogate. In many, but not all, cases such separation seems to have had serious consequences for adult personality, shifting it toward chronic mistrust, hostility, and delinquency.[8]

Experiments have recently shown that there are critical periods also for psychosomatic effects. It is well known that petting infant rats causes them to grow larger than control litter mates, and also to be superior as adults in certain avoidance-learning tasks. Similar petting at a slightly later age does not produce these effects. At first it was thought that the effects of handling were analogous to those of expressions of parental affection. This turned out to be a far too flattering misconception of the rat's reaction to being picked up by man. Electric shocks were found to produce similar effects. This and other studies showed that the effects were stressful, involving the links from the hypothalamus to the pituitary to the adrenal that we have already described.[40]

Furthermore, biochemical studies of adult animals showed that the brief handling during an early critical period had a permanent effect on the adult animal's physiological reactivity to stress, for example, the production of corticosteroids.

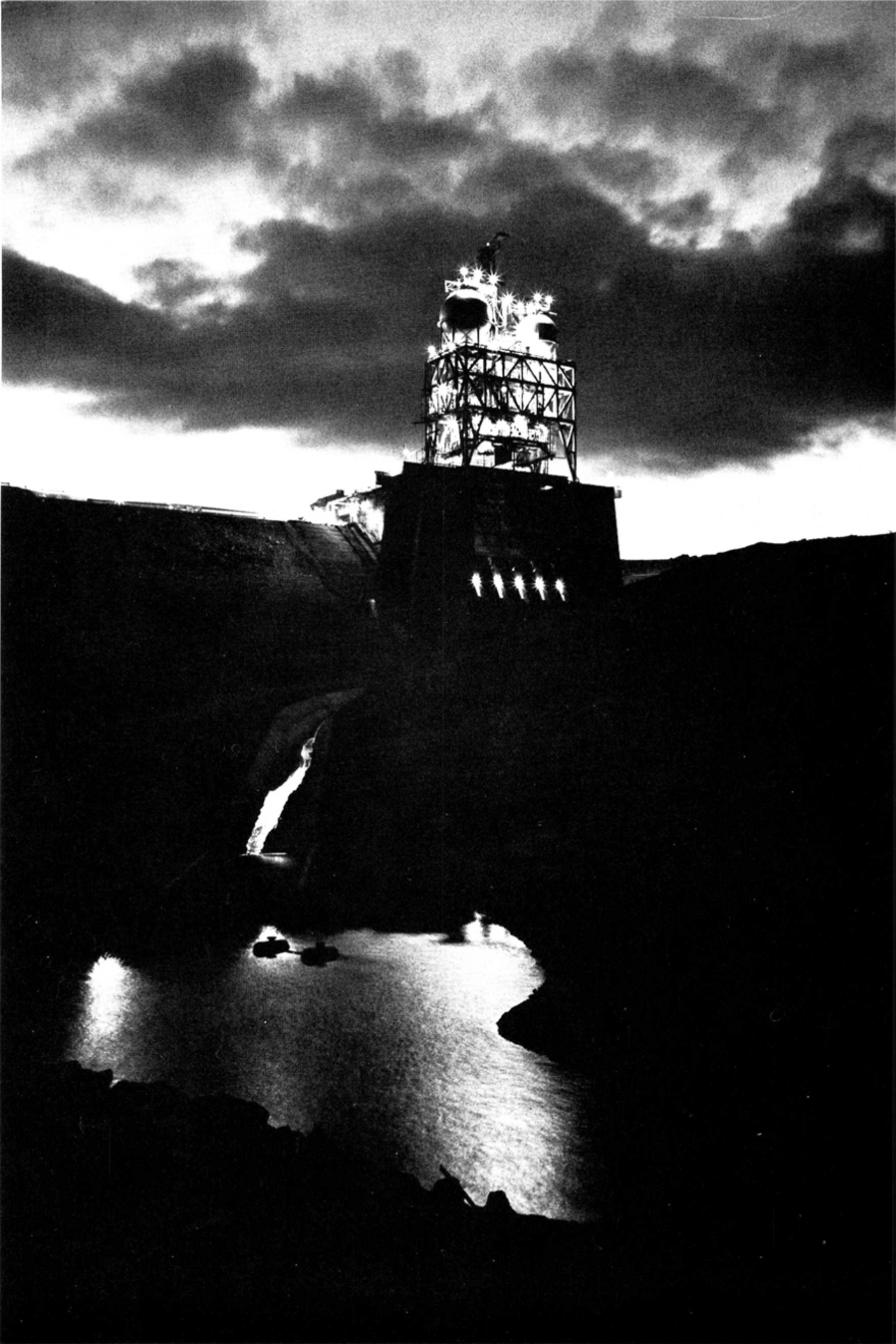
Additional studies are showing that there are critical periods of development during which the injection of certain hormones in minute amounts, which would be ineffective if injected at other times, can have permanent effects upon the rat's physiological development, which in turn have profound effects on behavior.[30] Such studies may help us to understand some of the physiological mechanisms involved in critical periods in development.

To return to our previous theme, do the psychosomatic effects of infant stress upon subsequent growth and maturation apply to human as well as to rat infants? One cannot perform experiments on stress with human infants. There are, however, certain societies in which infants at different ages are subjected to certain stressful procedures, such as cutting or burning the skin to form a pattern of scars, piercing the lips or the ears, or molding certain soft bones by pressure. Landauer and Whiting[38] have recently made a cross-cultural study of the effects of such procedures on average adult height. They have found that there is a statistically reliable positive correlation between such procedures and mean height. The average superiority in height of the stressed group is more than two inches, and the data suggest that the first two years are the critical period for children.

Since this is a correlational study, they cannot be certain that there is a causal relationship. Nevertheless, the correlation remains when they analyze the data in such a way as to control for three factors likely to influence height: namely, genetic stock, diet as indicated by protein, and sunshine, which presumably is related to growth-inducing vitamin D.

In short, clinical observations, behavioral experiments on animals, the results of biochemical assays and interventions, and cross-cultural comparisons are converging to show that there are critical periods in infancy during which certain events may have an especially great effect on subsequent physical and behavioral development. The phenomena are real; they are significant; the next step is to discover more about the laws governing them.

In conclusion, current research emphasizes the extreme importance of both innate physiological and acquired cultural factors in human development. We are beginning to learn a few of the fascinating details of the ways in which physiological nature and cultural nurture interact to produce behavior. But the picture is not a simple one, and each new discovery discloses additional tantalizing problems, challenging us to advance scientific knowledge deeper into the vast unknown.



The Scientific Endeavor

FOURTH SCIENTIFIC SESSION

GEORGE B. KISTIAKOWSKY, Chairman

Communication and Comprehension of Scientific Knowledge

ROBERT OPPENHEIMER

The Role of Science in Universities, Government, and Industry: Science and Public Policy

JEROME B. WIESNER

Synthesis and Applications of Scientific Knowledge for Human Use

J. B. FISK

Science in the Satisfaction of Human Aspiration

I. I. RABI

Power plant for launching Saturn V rocket to the moon, under globular fuel and oxygen tanks on test stand at Edwards Air Force Base in California. Each F-1 engine mounted on the stand generates 1-½ million pounds of thrust. [*preceding photograph*]

Introduction

GEORGE B. KISTIAKOWSKY, Chairman

BILLIONS OF YEARS ago living matter profoundly though unwittingly altered its environment and thus made possible evolution of life into higher forms. As Dr. Wald so eloquently explained in the session on the History of the Universe, billions of generations of millions of species participated then in the conversion of the earth's atmosphere from a reducing into an oxidizing medium.

We are living in an age when, for the second time in the history of the earth, living matter can change its environment on a global scale. This time it is a single species, *Homo sapiens,* in a period of time measured in tens rather than billions of generations, which is capable of altering the earth's environment. Weather modification, global management of fresh water and of vegetation, utilization of mineral and energy resources, and manipulation of continents' topography are powers which man possesses or will soon possess because of his knowledge of nature.

The urge to understand has been a salient characteristic of man since the beginnings of recorded history. Because of his unique ability to transmit experience and accumulate knowledge, which was stressed by Dr. Mayr in the third session on The Determinants and Evolution of Life, man has come into a position of dominance in the living world. The scientist's role in this process earned him a public position next to the philosopher, artist, and poet, until after the time of the Renaissance when a new tool was perfected by the use of which he could project a new, more exclusive public image. This tool was the experimental method which led to the flourishing of natural science in the Western World.

Our Academy's birth occurred at a time when systematic use of accumulated scientific knowledge for practical purposes was first being made, which soon caused the flourishing of synthetic chemicals and metallurgical industries. Only recently did societies as a whole, through governments, recognize science as a powerful stimulus for political, economic, and cultural change and progress; and thus only recently did governments begin consciously to promote science. The result has been its continuing exponential growth, the appearance of many new

GEORGE B. KISTIAKOWSKY Department of Chemistry, Harvard University

scientific disciplines, and also the emergence of many difficult problems. President Kennedy referred to some of the problems in his address at our Centennial Convocation. Some others are considered by the distinguished contributors to this symposium.

The extraordinarily rapid growth of science (let us not forget that over 90 per cent of all natural scientists who ever lived are still alive) and the proliferation of new areas of research, have led, from sheer volume as well as other factors, to difficulties in communication. There is the problem of transmitting information among the sciences, to preserve their coherence if not unity; an individual scientist knows more and more about less and less. A far more difficult problem is communication between scientists and those who apply the results of scientific research to political, technological, and cultural purposes. This is the topic about which we shall hear from Dr. Oppenheimer.

Dr. Wiesner will discuss our Government's effort to expand science in the expectation of rich practical returns for our society and to enhance its cultural contributions.

Dr. Fisk will tell us of the problems which arise when knowledge obtained in basic research is being translated into practical uses by organized modern industrial research teams. These are the uses that constitute the progress of our technological civilization.

And finally Dr. Rabi will speak of science in the framework of the culture of a free society as a force of far deeper significance than mere satisfaction of material needs.

✲

Communication and Comprehension of Scientific Knowledge

ROBERT OPPENHEIMER

THE THEME THAT has been assigned to me seems in some ways a little odd. That is only in part because this talk comes after three days and fifteen lectures in which, as actors and auditors, we have lived with many beautiful examples of good communication, and even very largely good comprehension — good understanding — of scientific knowledge. If I have any doubts, it may be that here and there, in those reports which dealt with subjects close to me, the communication and the understanding have gone a little bit beyond the knowledge.

In an important sense, the sciences have solved the problem of communicating within and with one another more completely than has any human enterprise. I may retell an old story. Thirty-five years ago, Dirac and I were in Göttingen. He was making the quantum theory of radiation, and I was a student. He learned that I sometimes wrote a poem, and he took me to task, saying, "In physics we try to say things that no one knew before in a way that everyone can understand, whereas in poetry. . . ."

It is an old and consistent tradition with us to be concerned with the words we use, and with their purification, and thus with the concepts in terms of which we describe nature. It was true of Newton, of Lavoisier, of Cauchy, of Mendel, and of course, in our day, of Einstein and of Bohr. As for Newton, we will understand this better when we have, after almost three centuries, the critical edition of the *Principia*; at least we will know that in the renowned "Hypotheses non fingo" it is not the first word but the last that bears the meaning.

When we tell about our work, we explain what we have done and we tell what we have seen, whether we are describing a radioastronomical object, or a new property of fiber bundles, or the behavior of men attempting to solve problems. We are prepared to believe that the explicit content of science has its roots in

ROBERT OPPENHEIMER Institute for Advanced Study, Princeton, New Jersey

these accounts of action, often factual, often foreshortened and synoptic, because cast in terms which the scientific traditions have established long ago.

Among us there is surely a great and appropriate variation in how we describe this foundation for the objectivity of our knowledge, and for the lack of ambiguity in the terms we use to tell of it; and of course there is an even wider latitude, in so far as we may bring ourselves to speak of them, in what we think of the reasons for the success of science, in what attributes of the world of nature in which we find ourselves underlie the manifestations of order which are our business: why we can work on the same table and with the same test tube when we cannot have the same melancholy or the same resolution; why so much of the order of the natural world finds its expression in number and the more abstract mathematical structure.

We probably all, with varying enthusiasm, would say "yes" to Charles Peirce as to "how to make our ideas clear." We would make a good case that we do indeed know the structure of some ribonucleic acids, or some properties of the longer-lived particles of physics, only leaving room for the fact that in new things as well as in old, there are points we may not have looked at, and that wonders may be hidden in the crevasses.

This foundation for knowledge precludes much that is an essential part of man's life. One cannot be a very effective scientist if he is a practicing solipsist. We cannot expect to describe a common world of introspection by telling people what we have done and what we have seen; though probably we can, and increasingly we will, describe elements of behavior which may have some correspondence to the inner world. Among these things of which we cannot talk without some ambiguity, and in which the objective structure of the sciences will play what is often a very minor part, but sometimes an essential one, are many questions which are not private, which are common questions, and public ones: the arts, the good life, the good society. There is to my view no reason why we should come to these with a greater consensus or a greater sense of valid relevant experience than any other profession. They need reason, and they need a preoccupation with consistency; but only in so far as the scientist's life has analogies with the artist's — and in important ways it does — only in so far as the scientist's life is in some way a good life, and his society a good society, have we any professional credentials to enter these discussions, and not primarily because of the objectivity of our communication and our knowledge. But if I doubt whether we have a special qualification for these matters, I doubt even more that our professional practices should disqualify us, or that we should lose interest and heart in preoccupations which have ennobled and purified men throughout history, and for which the world has great need today. Your lives attest this.

This account of a constant concern within the scientific enterprise to purify and refine our language is, of course, a sort of parody of what we are all about. We do not really do this except in moments of crisis, or in order to make way for something very new and deep. We come to our new problems full of old ideas and old words, not only the inevitable words of daily life, but those which experience has shown fruitful over the years. This is an inevitable approach to the new; and when it is not too new, it gets by. But the comprehension, the understanding of scientific knowledge is a very different thing from being the recipient of a communication. I think there is an element of action inseparable from understanding: to question, to try, to apply, to adapt, to ask new questions, to see if one understands, and to test what one has been told: action in the laboratory or the observatory, or on paper, or, at the very least, in the motions of the spirit. We need, at times, to talk about the sources and the springs of this motion, without which communication would provide the fuel pipes, the electrical wiring, the transmission of a car, but not the combustion which gives it power and life.

We do not talk of this very well; imagination, play, curiosity, invention, action, these are all involved. They are indeed only rarely all combined, and supplemented by skepticism and criticism, in any one man in any one moment; one of the charms of the scientific enterprise is how deficient we can be in many of these qualities and still play some meaningful part in it.

We know that we love the old words, the old imagery, and the old analogies; and that we keep them for more and more unfamiliar and more and more unrecognizable things. Think of "wave," "information," "relativity." We know that one can explore and study the springs of the movement of science, that it is a fit if very difficult subject of study. Today at least we are not able to talk about it very well, not at all as we can of molecules or galaxies, or even of the effective definition of the words that we use. Yet we may be sure that without a living engagement there is no understanding and there is no life of science, and we know that we cannot command this, or perhaps even learn it, except by apprenticeship, by following what others have done, and by listening to the mischievous voices of adventure and play and exploration and doubt with which we greet a new experience or a new communication. This has very much to do with what we can in practice and honestly mean by the unity of science. I think, for instance, of contemporary mathematics, whose absence from this program does not at all reflect a lack of vitality, of discovery, and of beauty in the current scene. Up to our time, it has been the experience of our enterprise that there have been a good number of men who combined creation and wide knowledge of the mathematics of their day with a lively interest in those elements of the natural sciences in which this mathematical order might be embodied. This conversation, as a lively mutual understanding, is rather thin today. It is not rare to find a physical scientist who will hear some beautiful new result—in algebra, for instance, or topology—with pleasure, with amazement, and with admiration; but it is not likely that he will be deeply engaged, and try to see if he can make it wider, how it affects other things he may have known, or thought to know. I know that it is also true that many mathematicians will accept with a certain interest that there are in nature two neutrinos which have different properties, or that astronomers believe that they may be witnessing evidence of very massive gravitational implosion in other galaxies. To me it seems good that we still do tell each other these pieces of news; but I would hope that the century-long tradition of a felt sense of reciprocal relevance between mathematics and natural science would soon again find itself embodied in many of us, or, far more plausibly, in our successors.

Thus between us, as specialists in our professions, there is a partly accidental quality to the effectiveness of our converse with one another, and thus to the effective unity of our view of the world, even as scientists. There are two reasons for deploring this. One is that past experience suggests so strongly that among the sciences there are elements of relevance and mutual enlightenment which make such converse an essential part of deep and rapid progress; the other is that we regret for ourselves what we do not really know, and we regret for others what we cannot really tell them. This is, of course, a reflection, within the internal society of the scientific enterprise, of a situation that characterizes our relations with human societies as a whole, with the society within which we are embedded, and that leaves us with problems, some very grave, and by no means all clearly soluble, having to do just with the communication and comprehension — understanding — of scientific knowledge.

These problems rest, of course, on human weakness and limitation; but more specifically they rest on at least three features of the scientific enterprise which it has in common with the world in which its whole action takes place: size and saturation, growth and change, and specialization. I will not speak to size, having no true wisdom as to whether there is a natural and appropriate limit to how vast the scientific enterprise can and should be, beyond which it suffers too deeply from suffocation and fragmentation. I do not think that I know an answer, but I rather hope that those who follow me in this symposium may have some wisdom; and I know that they have some views on how large our world should be, perhaps because that is a still harder question.

One thing we do know: growth and change imply size, and growth and change are very deep in the nature of the scientific enterprise. Without them we would not recognize the rooms in which we were living, or what our days were all about. As for specialization, it is what sharpens our tools and our words, and is the instrument for penetrating deeper and farther into the world of nature.

I think that we must live with these, and that we can live with them. Some of us will know one thing, some fewer will know many things, and the unity of our knowledge, its freedom from contradiction, and its important and often very deep common relevance, will not preclude but will be enriched by the great and blessed diversity of man.

These limits on the communication and comprehension of scientific knowledge which we find among ourselves, with which we have been living and will continue to live, have their analogies in the related but vaster limitations that we have in our external relations with those who are not yet, or not ever, involved in the scientific enterprise. The first of these is with the young, those who may be entering the life of science, and perhaps also, perhaps even more importantly, those who may not. I cannot speak with even a decent record of experience or authority of the problems of education and schooling, for I have known them only at the late level that is essentially apprenticeship, where a young man or woman has become engaged in some part of science, and the problem is to help him enlarge his interest and his power and his knowledge of what others have done. I have the impression, which I hope may be true enough to be shared by most of you, that in the graduate schools, and in their increasing postdoctoral studies, we have in the natural and mathematical sciences rather happy arrangements for this period of apprenticeship, happy in comparison with the situation in other branches of study — historical or philosophical, for instance — rather happy in comparison with our sister institutions abroad, and very happy indeed in comparison with our own country some fifty years ago. Apart from this, my life gives me no qualification except to express an appreciation to our many colleagues who have been studying and practicing the teaching of the sciences in the schools and the colleges, so that first sight shall not repel, and the institutions not resist the natural curiosity and love and joy of the experience, but open it, so that as many as can will have an opportunity to discover some trait of nature, to see with welcome some sure sign of order in nature, with their own hands and their own heads.

I know that our colleagues understand the universal value, in all teaching, of quickly correcting error. I know that they are concerned to free the teaching of science of a slovenly and lazy dependence on history, in which discoveries were often made in obscure, contingent, and not deeply relevant struggles, whose interest as history is not helpful to the young student, and usually obscure to his teacher. I know that they hope, as often as may be, to open perspectives on the larger connections in nature and in the sciences which describe it, and rarely, when it can be done with historical scruple, on some chapters in the history of man's knowledge.

We mostly take it for granted, though it is not quite obvious, that we would like to have this opening of the world of science, this induction into it, effective not only for those who will be of our company, but for as many as may be of all the young, and the newly young who are willing to study. It seems to me that there are probably two reasons why we hold this view, not in the first instance commensurable reasons. On the one hand I think we increasingly feel the need for companionship and for help. I am not here speaking of the patronage of science, which has not been ungenerous or niggardly in the past years, though it may come to be so. I have in mind rather what we all know, that more rapidly than ever before, the sciences have been embodied in new technologies, and that these bring on the scene new powers and possibilities, now a new need, now a new opportunity. These needs and opportunities often are relevant to what in us, and in most men, are the most deeply held convictions of what is right and good, convictions rooted in a long tradition, and integrally a part of our sensibility. We do not talk about it much, but most of us, I think, are committed to preserving life and health where that is possible. Increasingly, and largely because of the effects during the last centuries of technology and industrialization themselves on its modes, we are committed to limiting, if possible to eliminating, war. We are committed to relieving, to reducing labor and drudgery, and not only the hard labor of the field and the mine and the galley, but the dull labor of the Midlands factory. We are now clearly engaged in a great enterprise testing whether we can live in a world in which war does not play its traditional part, an enterprise in which not only long-inherited human institutions, but even older, even other more permanent human attitudes, of anger, hatred, solidarity, self-importance, righteousness, which war has fed, can permit the change. We are in this too deeply, I think, to let the good news or the bad news of the day or month or year affect or limit our hope and, where it is possible for us, our engagement in this great, open, unsettled action of man's history.

With the preservation of life too, and along with it the alteration and automation of work, we are concerned not only with the inadequacy of our institutions, which were framed for a very different world, but with our attitude toward the meaning and value and nature and quality of human life, so largely in our past built on productive work as its foundation. Here in this country we see the mixed fruits of medical and engineering technology first with the young and the old. It is reasonable to expect that they will spread, and that they will characterize many other technologically developed societies. I know of the concern, so well expressed by some of you, that even the saving of children's lives may have created problems with which no one can cope, that have some bearing on the growth and size of the human society.

Though I do not suppose that a thorough knowledge of science, which is essentially unavailable to all of us, would really be helpful to our friends in other ways of life in acting with insight and courage in the contemporary world, it would perhaps be good if in talking with them we could count on a greater recognition of the quality of our certitudes, where we are dealing with scientific knowledge that really exists, and the corresponding quality of hesitancy and doubt when we are assessing the probable course of events, the way in which men will choose and act, to ignore or to apply, or make hypertrophic or nugatory the technological possibilities recently opened. I think that some honest and remembered experience of the exploration of nature, of discovery, and of the way in which we talk to one another about these things, might indeed be helpful; but that is because it would remove barriers and encourage an effective and trusting converse between us, and make more fruitful the indispensable role of friendship. These things are perhaps always easier in a small society. They were perhaps easier a century ago, for us and for many of the countries of Europe. We have a modest part to play in history, and the barriers between us and the men of affairs, the statesmen, the artists, the lawyers, with whom we should be talking, could perhaps be markedly reduced if more of them knew a little of what we were up to, knew it with pleasure and some confidence; and if we were prepared to recognize both the important analogies between what moves us to act and to know, and the extraordinary and special quality of our experience and our communication about it with one another. I have often thought that with the historic game so grand and so uncertain, we should not dismiss any help, even of that small part which we could play.

The other set of reasons for hoping that young people who will not be professional scientists, and older people who are young in heart, could have a greater scientific literacy and some limited experience, as ours also is limited, is that we know, all of us, that the experience of scientific discovery is a good and beautiful experience, and an unforgettable one. We know that this is true even of little discoveries, and we understand that with the great ones it is shattering. It was on his seventy-first birthday that Einstein said to me, "When it has once been given a man to do some sensible things, afterwards his life is a little different." It seems not really as an act of arrogance but simply human, and not in the purely pejorative sense of the word, to wish these pleasures for as many of our fellows as can have them.

In our world, many things that men do rather naturally, that they have learned to do long, long ago, have become professions, have become part of the market. I think of song and sport and the arts, the practical arts and the fine arts. None of these is without discipline; and although they are very different from those that lead to the sciences, I would be slow to rate them easier. Yet people sing and make sport and practice the arts quite apart from the market, quite apart from a career. It would be a poorer, thinner life without that. Though surely we will not all burst into song, or take to skis, or pick up a chisel or a brush, some of us have done some of these things, and some of us will; and it seems a proper hope that in our education, both for the young, and for those, in growing number, who like us have kept a lifelong taste for it, we do what we can to open the life of science at least as wide as that of song and the arts. Not everybody will want our pleasures, as among us not everyone can taste the other's, and as even we cannot expect an astronomer and a biologist fully to share what each has. We think of this as a high and lovely part of life which, with all its discipline, is still directly responsive to a deep human need. We all know this, and all share it; but each of us, I think, must be free to use his own words to sing its praise, even to describe it.

We may be seeing a time in which war will come to play a smaller and an increasingly trivial part in man's life. I hope that we are. We may be coming to a time in which for growing parts of the world the production of goods will require a much more minor commitment of human effort and life, and the market leave men with a far greater measure of freedom. I hope that we are. For this it will clearly not be enough that we preserve the integrity of our communication and comprehension, either among us, or with our fellows; but this is at the heart of our enterprise, and it is the least we can do.

The Role of Science in Universities, Government, and Industry: Science and Public Policy

JEROME B. WIESNER

THE ANNOUNCED title of my talk was provided for me by Dr. Bronk, who was intent upon making my task just a little bit easier. While I have cast in the mold that he provided, a more abbreviated title would be "Science in the Affluent Society."

We are here at this time to celebrate the centennial of our Academy, to honor its founders and the many men who have added to its luster since then. In the previous meetings of this celebration we have listened with excitement to reviews of the many fields of knowledge that compose the body of science. Now it is our purpose to consider the many other faces of science, to review and assess our

JEROME B. WIESNER Special Assistant to the President for Science and Technology, The White House, Washington, D. C.

estate and look ahead to the opportunities and needs in science and in the society of which we are a part.

In recent years scientific and technological achievements have been a matter of great public interest as have the many unanticipated problems which the widespread exploitation of scientific knowledge has produced.

It should be gratifying to find in this centennial year that the general interest in science is greater than ever before. In the newspapers and other publications, in the Congress too, one finds a growing interest in research and development. There are at the moment several Congressional committees examining the purposes and methods of government-financed scientific and technological activities. While such interest is not new, the point of view seems new. In the recent past, discussions of science often consisted of uncritical praise. Now the situation is changed. Serious questions are being asked and many of them reflect deep-seated concern about the character and purposes of the nation's scientific and technological undertakings. Many also reflect a desire to become more familiar with these processes.

Why has the mood changed so? Has science changed, or was too much expected from science in the past, or have the nation's needs changed? I don't believe that the answer is really in this direction. What has changed, I believe, is the breadth of motivation for supporting research through the Federal Government. There has been a broadening of emphasis from a primary need to support military development to a wider purpose encompassing the entire spectrum of social needs. The military objective is of great significance, to be sure, but now this is only one of numerous reasons for many of the new research and development activities.

To be specific, while the level of spending for military research and development has remained almost constant during the past 3 years, total federal expenditures for all research and development have continued to rise at an exponential rate. This exponential has a doubling time of about 4 to 5 years, and the last factor of two involved an increment of approximately \$7 billion. Military research and development is still the largest single component of the federal research and development budget amounting to approximately \$8 billion of the \$14.5-billion budget for fiscal year 1963.

When the increases in the research and development budget were primarily for improved military security, they were easier to understand than they are today. In retrospect, our nation was very fortunate — I hope this is not misunderstood — that this incentive for research support did exist following World War II. I doubt that there existed at that time a sufficient appreciation of the general importance of scientific research to have made possible the creation of our present large and very competent scientific establishment on the basis of needs for the general welfare.

Recently in reviewing the political scene, the President observed that there are cycles in the affairs of men. And in science, too, we can observe this fact. We have experienced a heady period of growth and the time has arrived to review objectives, assess accomplishments, and adjust stressed institutional arrangements to present needs.

We have been through these periods of growth and integration, analysis, and reorientation before — many times — during our history as a nation. Starting with the Constitutional Convention, the American people have had a concern for science. And it has always presented a problem to their government. At every period in our history, science has contributed to the development of the nation, and at every step needs of the Federal Government and federal funds have been a major factor in achievements of science. At some point in every historical period the question of the role of the Federal Government in science and the search for the proper form of science organization within the Federal establishment have been burning issues. Though the research efforts involved may appear minuscule by our standards, they were as important to their times as our much bigger programs are to ours. Physical surveys, exploration, standards of time and measure, geological surveys, patent incentives to spur invention, navigation, agricultural science, and public health were among the early issues that involved the government in scientific endeavors.

In his book *Science in the Federal Government,* A. Hunter Dupree traces the evolution of scientific organizations in the government, and here one can clearly see the ebb and flow of enthusiasm for research and technology with passing needs and opportunities.

Most interesting to the contemporary scene is the violent debate which took place over a Department of Science in the 1880's, an argument that was resolved by the Allison Commission, a joint Congressional Commission, when it concluded that the government's scientific establishment and the scientific community in the universities had already grown too complex for such a change in organizational structure.

It is interesting to note that the recommendation for a Department of Science came from a committee of the National Academy of Sciences. It thus appears that committee work for scientists is not a new invention, nor for that matter are government consultants who figure prominently in the early history of Federal science.

One important characteristic stands out above all others in the development of science in the Federal Government, and that is continued growth. In periods of national emergency and in periods of enthusiasm for science, the rate of growth has been faster — during periods of reassessment or preoccupation with other problems such as the great depressions, it has been slower but the averaged slope continues upward. I have examined manpower data which goes back about 40 years, and the steady rate of growth can be seen in them. During this interval, the number of scientists and engineers has doubled every 12 years. If one assumes that the same rate of growth has persisted since 1800, there would have had to be about 100 individuals engaged in scientific and technical activities, a not too unreasonable number.

I don't cite this history to absolve us of the need to deal with our current problems, but rather to give us some perspective and to indicate that any solutions which we devise today will undoubtedly not be adequate a few years hence. In fact, one need only review the history of the innovations introduced during the past 5 years to see how rapidly the problems change.

The post of Special Assistant to the President for Science and Technology, which I now occupy, was created in 1957. At the same time the President's Science Advisory Committee was made a part of the White House proper. Prior to that time it had been in the Office of Defense Mobilization.

The Soviet Union's launching of the first Sputnik provided the immediate impulse for these moves, but they were also a belated recognition of the greatly increased importance of technical matters in issues of national policy. Dr. Killian, as the first Presidential adviser on scientific matters, rapidly became involved in matters of the greatest national importance involving education, defense, disarmament, space, and international cooperation. In fact, I don't think it is stretching a point to say that the impartial assistance provided by Dr. Killian and the Science Advisory Committee made it possible for the President to arrive at many policy decisions which would have been impossible otherwise.

The creation of the post of Special Assistant provided a means for matching the activities of the Committee to the needs of the President. The full-time

Special Assistant was useful to the President in many other ways as well. As a regular attendee at National Security Council meetings, he was in a position to be alert for scientific aspects of a broad range of questions and to provide the President with a technically knowledgeable assistant, having a loyalty to the President alone, rather than to one of the operating agencies, to complement the scientific advice that he received directly from the agencies. During the first year or two, military and space-related problems were of such urgency that they commanded nearly all the attention of the Special Assistant, his staff, and the Committee. In retrospect this was quite appropriate, since the new arrangements were a response to a failure to fully appreciate or adequately manage a revolutionary technical-military situation. It is clear now — looking back — that the period 1950 to 1960 was truly revolutionary measured in terms of speed, power, and ability to communicate: the three parameters that dominate military capability. The atomic bomb had multiplied deliverable explosive power by a factor of a thousand and the thermonuclear weapon multiplied this by another thousand. The ballistic missile reduced the reaction time of intercontinental combat from tens of hours to tens of minutes. And the electronic computer made it possible to replace man, whose reaction times are measured in tenths of a second, with automatic devices thousands of times faster. Sputnik resulted in other important organizational modifications; in particular, the Office of the Director of Defense Research and Engineering was established in the DOD and a civilian space agency was created to provide a focus for the peaceful exploitation of space. The changes in the Defense Department management structure established a scientist or technologist in the top management of the DOD and contributed much to the rationalization of the technical activities which has occurred in that department.

It is interesting to note in passing that President Eisenhower early turned to PSAC for assistance in unraveling the conflicting information he was receiving regarding the technical aspects of the nuclear test ban which he was attempting to negotiate with the Soviet Union. Starting in April of 1958, the members of PSAC became the President's major source of technical assistance in his attempts to find a mutually satisfactory safeguarded agreement. It was not until the Arms Control and Disarmament Agency was created by this Administration that a continuing full-time effort existed in the Federal establishment to understand the many complicated interacting technical, political, and military issues that must be taken into account when considering an arms limitation or disarmament step.

The Science Advisory Committee also assisted President Eisenhower with technical issues involved in more comprehensive disarmament efforts, and several members of the group, including Dr. Kistiakowsky and myself, actually participated in negotiations.

For the first year and a half or so, the Special Assistant and the Committee devoted most of their efforts to security issues, but by 1959 the growth of the research and development programs in the agencies was sufficient to require some coordination and integration and another step was taken in the formation of the Federal Council for Science and Technology. It was composed of policy officials of the principal agencies engaged in research and development. The Council was to function as a sub-cabinet for science and to provide advice to the President concerning governmental scientific activities with particular emphasis on problems affecting more than one agency, and to make specific recommendations for the planning, administration, and coordination of federal programs. The Special Assistant also acted as chairman of this group.

When I came to Washington in 1961, I encountered considerable pressure for the creation of a stronger coordination and science policy mechanism at the Presidential level. In fact, there were, and I might add still are, strong proponents of a department of science, complete with Cabinet-level secretary, both in the Congress and among members of the scientific community. The Science Advisory Committee studied the problem for approximately a year and concluded that a department of science, bringing together many research activities now found in the individual agencies of the Government, was not advantageous, but that a need did exist for a more effective and more comprehensive oversight of the federal scientific and technological activities.

Last year an additional administrative mechanism was added to meet this need. The Office of Science and Technology was established by a Reorganization Plan that in effect institutionalizes and makes more permanent those staff resources that had been previously established to advise and assist the President.

Possibly the most important consequence of providing a statutory basis for the scientific activities in the Executive Office of the President is the fact that it enables the Director to appear before the Congress to explain, when possible, the government-wide views of issues, activities, and problems. During the past year I have testified before a great number of Congressional committees and have found this an enjoyable and useful experience. I hope that the committee members share my reaction.

Not many years ago in his provocative book entitled *The Affluent Society,* Kenneth Galbraith illuminated the new problems in transition when he attributed many of the economic and social ills of our nation to our failure to appreciate the fundamental changes that technology had made in our industrial society. He pointed out that modern machinery and automation had made it possible to produce all the consumer goods which our people would ever desire to consume, using a fraction of our total labor force; that following the conventional wisdom existing among economists, we still practiced an economy of scarcity in a society where scarcities need not exist. To use his words:

> The final problem of the productive society is what it produces. This manifests itself in an implacable tendency to provide an opulent supply of some things and a niggardly yield of others. This disparity carries to the point where it is a cause of social discomfort and social unhealth. The line which divides our area of wealth from our area of poverty is roughly that which divides privately produced and marketed goods and services from publicly rendered services. Our wealth in the first is not only in startling contrast with the meagerness of the latter, but our wealth in privately produced goods is, to a marked degree, the cause of crisis in the supply of public services. For we have failed to see the importance, indeed the urgent need, of maintaining a balance between the two.

In this eloquent prose, Mr. Galbraith was pointing to our present dilemma: how to establish a proper balance between satisfaction of our individual wants and the necessary or desirable needs of the community as a whole. There is obviously no simple, or even single, answer to the question of how to best use our available resources. Public support of science has until now largely been exempted from this debate because funds for it were a small part of very large sums of money provided to ensure our military security. The only substantial sums provided for another purpose in recent times were those related to health needs. The unfulfilled social and economic needs around us coupled with one other fact — a leveling off of research and development in relation to gross national product —suggest that the nation faces a significant era of re-evaluation. This process of leveling off explains the reason for many of the current questions. It indicates, as well, that confusions which were of no consequence 10 years ago must be gotten straight today. To the extent that federally supported research and development is justified for social purposes other than national security, it will be judged by other standards, less well defined, and more controversial as well. I would like to call to your attention some of the matters of concern that I encounter most frequently.

There is concern about waste of funds and imbalances in the federal programs. There is concern about distortion of our federal activities and our universities. There is worry that the unanticipated but inevitable side effects of new technology are causing more and more difficult problems which then require further and more expensive remedial actions. Our agricultural surpluses, air and water pollution are good examples. The unpredictable consequences of the widespread effects of the use of pesticides is another, technological unemployment yet another. There is concern that federal research and development expenditures are responsible for unbalanced economic development between geographic areas of the country and between different industries. There is fear also that too large a fraction of the ablest youngsters are being attracted into science and engineering.

Underlying it all is the belief that the whole activity is beyond the comprehension of the individuals in the government who are responsible for it, be they in the Executive Branch of the Government or in the Congress.

Some of these worries stem from real problems, involve the need for policy actions, and should be considered and deliberated. Others stem from misunderstandings which we should strive to eliminate.

For example, research activities and development are frequently not distinguished. They are mixed together and called science. Even most engineers and scientists fail to make a clear enough distinction between activities carried out solely for the purpose of adding to existing scientific knowledge and work that is performed because it may be able to satisfy some practical need. And this confusion is at the root of much misunderstanding among non-scientists who are called upon to make decisions or pass judgments about technical matters. It isn't that most people fail to recognize that there is a spectrum of activities with pure research at one end and hardware development at the other. The problem is rather a failure to understand that the methods and motivations of each are different, and that often research, new knowledge, is necessary in order to achieve a practical goal.

Not only are research activities quite different from development work, but so are their costs. We have estimated that the fiscal-year 1963 obligations for basic research by the Federal Government are approximately \$1.4 billion out of the approximately \$14.5 billion devoted to research and development and of this \$1.4 billion one-half billion are funds for space science and include the cost of boosters and launching operations. I quoted this figure during a recent Congressional hearing and was later told by Congressman Pucinski that he hadn't realized that basic research was so small a portion of the total. "Maybe, "he said, "they were focusing on the wrong problem when they focused their inquiry on basic research expenditures rather than on those for development."

In my view development activities, the creation of useful new devices, should only be undertaken if there is a clear-cut requirement for a new product after it has been developed. This is reasonable, for it is ordinarily possible to make satisfactory predictions about the probable cost and performance of a proposed new device, be it an aircraft, a computer, a chemical processing apparatus, or a nuclear power plant. Consequently, it is also possible to make a decision about the desirability of a given development. Furthermore, because development efforts are generally much more costly than research, one should apply rigorous tests of need before starting new efforts.

In the case of exploratory development and applied research, there is reason to be more venturesome. Here the search is to see whether practical applications of new knowledge are possible. For example, there is a vast and varied effort today to explore possible new uses of the laser, the source of coherent light developed recently, and elements of this work should be carried to the point of demonstrating the feasibility of underlying concepts. Here, too, work beyond such a point should be permitted only if the ultimate capability is needed. These criteria are applicable to all but the basic research segment of the research and development effort. In other words, a basis exists by which administrators and legislators could establish the need for more than 90 per cent of the federal expenditures for research and development. I am not saying that they could or should pass on the technical validity of proposals or judge between competing ways of accomplishing a given objective. This will still be the function of experts, but then the experts will be passing on means, not ends. Corporation executives, government budget officers, and department heads and members of Congress have traditionally made such decisions with confidence, and with access to scientific and engineering advice; there is no reason why they cannot continue successfully.

But the choices in the field of basic research must be left to the scientists. This is why I place so much emphasis on the matter of distinctions. Even here, others will need to make decisions regarding the over-all level of effort, and if that level is less than is required to support all the worthwhile research that scientists want to do at any given time, it will be necessary for the scientists to make decisions regarding the support for the different disciplines. Not that this would be easy either. I could write an entertaining book about our efforts to make decisions in the field of high-energy accelerators.

I am very frequently asked whether there are no limits to the growth of research activities, and I find it impossible to give a simple answer to the question. There is no foreseeable limit to the amount of productive scientific research that can be done. On the contrary, I believe that Vannevar Bush was making an understatement when he called science an "endless frontier." It is an expanding frontier as well, if that is conceivable, for on the average each working scientist raises more questions than he answers.

Clearly, the lack of interesting and important work will not set a limit to the extent of our research activities. Many more people have the potential to do effective, creative research than are currently engaged in such work or are studying to become scientists. I will say more about this later. It appears then that the amount of research will be set by the willingness of the country to pay for it. Though we have not passed the point where we can fail to reap increased benefits from increased expenditures on basic reaserch, we have reached the point in time where the rate of growth for the total research and development effort will probably be diminished. I should report a very interesting fact encountered as we prepared for my testimony before the Daddario Committee. Though the actual level of federal research and development obligations rose from $11.2 billion in fiscal year 1962 to $14.5 billion in fiscal year 1963, the percentage of the gross national product for the given years represented by these figures remained constant at approximately 2 per cent. In this climate our basic research activities should be scrutinized with increasing care. Marginal activities should not be tolerated. Here considerable assistance is needed from the academic institutions and from individual scientists, too, for it would be extremely difficult and very undesirable for the government to set standards for academic research and even worse for it to try to police them.

Especially confusing to the lay observer, I find, are the interrelationships between institutions and individuals in the scientific community.

The American scientific enterprise today is a complex one, involving many different associations among the universities, industries, the government, and the individual scientists as well. It is a complicated organism that has very little defined structure. It is the product of evolution in a permissive atmosphere. A single entity in one group may have many relationships with subgroups in the other. For example, a department in a university, or even a single professor, may be engaged in research for several agencies of the government and possibly even for industry. Similarly, a single government agency may support many separate activities through independent and often quite different arrangements within a single institution. Finally, the faculty members of the academic institutions may be found consulting independently for both industry and government. All of this is mighty confusing when witnessed from the outside. There is little wonder that there are frequent proposals to rationalize the system. To be sure, improvement in procedures and more uniformity in practices among agencies are badly needed. But from the point of view of scientific accomplishment, there is great danger that much would be lost if highly structured arrangements were substituted for those we now have. The present arrangements have several important attributes. For the research worker there are a number of different agencies where he can seek support, both speeding up the process of finding support and making it considerably less likely that the prejudice of any single individual or group can dominate the research in a field. Second, it hastens communication, particularly between fundamental research groups and groups in industry or government whose work might be aided by the new knowledge. Third, the frequent contacts between groups result in an atmosphere of mutual understanding, respect, and confidence which are often lacking in situations where formal barriers inhibit them.

Many foreign countries do have much more formal arrangements for the support of research in their universities and industry, more rigidity in relationships between the different sectors of the society, and their scientific enterprise is less effective as a result.

I suspect that most of our colleagues visiting us from other countries would urge us to preserve the freedom and rapid communication our arrangements provide. If we are to preserve them, two things must be done. We must be sure that the great advantages of this pattern of operation are understood, and, even more important, that special care is taken by all the parties involved to ensure that abuses of the freedom thus provided do not occur.

From the point of view of the academic institutions, while federal support for research has made research programs stronger, other problems have been created. First of all, research activities are larger than they could be if the universities still depended primarily upon private philanthropy for support of research, as they did before World War II. Bigness, however, presents problems and many scientists deplore the fact that often modern research requires large expensive equipment with the concomitant need for team work and a high degree of organization to make effective use of such facilities. Laymen echo these laments without recognizing that many scientific problems today cannot be studied without such tools. Money hasn't created the problem in this case. In other research fields, the size of activities is related to the level of support, and in them it is true that government support has set the level. As we have already indicated, the motivation for doing this was to ensure that adequate basic research activities existed.

The government turned to the universities for increased basic research because universities traditionally have been an important source of new knowledge in the sciences. Such support was, of course, welcomed by the universities, particularly by the scientists, though many administrators were troubled by the long-term implications of this course. With this support, scientific research in the universities prospered and so did the scientific departments involved. Thus, this activity undertaken for the general public welfare proved very valuable to the academic institutions, too. There was a not wholly anticipated but vital by-product of these actions. The increased level of research made opportunities for more graduate students. If this had not occurred, we would be even shorter of highly trained scientists than we are today. In passing, I might note another popular misunderstanding, that faculty members when engaged in research are not participating in teaching. There is little general appreciation of the fact that the only effective way to train research workers is by the apprentice method. The need for scientific and technical graduates has been so great that the government through the Science Foundation and the National Defense Education Act has awarded fellowships and loans to further increase their number. Increased research activities have created demands for buildings to house the work, which demands have caused a serious drain on institutional resources. Provisions of the Administration's education bill were proposed to mitigate this problem. We are hopeful that the Senate will restore funds for this and other purposes eliminated from the National Science Foundation budget by the House and that the House will accept the educational facilities bill passed by the Senate.

One final difficulty for academic institutions stems from a view now current that an institution doing research with federal funds should make a contribution to the effort by sharing the overhead costs incurred by the work. This is one factor behind current efforts, some of them successful, to limit the overhead rate on research grants to a fixed percentage regardless of what the actual costs may be. Such limitations will save little money and may do much harm to the educational programs of our universities. This is not the only reason expressed for limiting overhead on academic research grants. I have encountered individuals who believed that overhead represented a profit, and, to be sure, a profit of 20 per cent on a government grant would be generous. Here again we see the seriousness of the lack of understanding of act or purpose.

As the scope of federally financed research activities increases, the problem of proper choice and supervision of grants becomes increasingly difficult. In some academic institutions government agencies must support and monitor hundreds of separate grants, most of them made to individual faculty members. This is a difficult and inefficient task, and the time has come to find a means by which the institutions can accept more of the responsibility for the allocation and use of federal research funds in their institutions. I suspect that neither the funding agencies, the universities, nor the individual researchers will welcome this suggestion, but in my opinion some such arrangement will be necessary if research activities continue to grow. This may prove to be the only possible alternative to the increasingly tight controls that are being imposed on investigators, and the almost impossible supervisory burdens that now fall on some agencies.

To me the most perplexing problem of all is that of reconciling our present basis for awarding research grants based solely on scientific quality and the need and desire to build up the academic and scientific excellence in many parts of the country where it has not previously existed. There is a widely held view that the present system discriminates against areas with modest scientific establishments; that the federal funds are used to attract scientists and students from such areas to the large centers and so, in effect, the rich get richer and the poor get poorer. With the growing realization that economic well-being will be ever more closely linked to technologically based industry, there is a growing determination in most areas of the country not already so endowed to create a strong scientific base. Hence the growing resentment and criticism of the manner in which the federal funds for research are allocated.

I do believe that it is in the nation's interest to provide assistance to academic institutions that have the clear potential for becoming centers of excellence in graduate education and research, particularly in geographical areas aspiring to acquire a technological base. Nonetheless, I believe that it would be disastrous to the scientific enterprise if standards other than quality were made the primary basis for the allocation of money for support of basic research.

For that reason, the Administration has asked the Congress to provide some special funds, in the Science Foundation budget and in the funds for the Area Redevelopment Authority, for building centers of research excellence. In this way discretion could be used in providing a base of funds to permit an institution to strengthen its facilities and competence with the objective of becoming competitive for the normal research funds. In this way it would be possible to help the deserving schools and needy areas without destroying the present effective mechanism.

Unfortunately, the House Appropriations Committee did not support this NSF budget proposal.

These issues and many others will undoubtedly be examined by the House committees investigating research and development. The future for all of science will be significantly affected by the outcome.

In this regard, many individuals, many organizations, will inevitably be parties to an inquiry that will be undertaken by the entire nation as well as for the entire nation. It is in this set of circumstances that the National Academy of Sciences could play a highly significant role. I expect that the National Academy Committee on Science and Public Policy will consider many of these issues, too, and their reports will be very important in the continuing discussions.

The next century of service to the nation by the Academy could begin on no more important note than its contribution to a broader comprehension by all sectors of our society — the scientific community included — of emerging implications and opportunities of science in our affluent society.

Synthesis and Applications of Scientific Knowledge for Human Use

J. B. FISK

SURVEYS OF THE "scientific endeavor" naturally touch on what has happened concerning the *applications* of science in the last century, especially here in the New World. It might be expected that, with our background of social and economic change and the impact of the industrial revolution which had come with full force from Europe about a century ago, we would find in the application of science in America in the past 100 years new patterns and experiences. These could be of special import for future thought and action now that we are in full passage through an Age of Science.

Indeed, there are things to be learned from this perspective of the past century, but much of what we shall now discuss began even earlier, especially in the modes of application of science that were foreseen in western Europe more than three centuries ago. Thus, it is especially timely in assessing this feature of the endeavor of science to see how far we have come and in what directions we have moved with respect to the great original visions of what the practical values of science and technology might be.

The keenest perspective on this is found in the writings of Francis Bacon. Bacon's doctrines were in turn interpreted for a later period by Lord Macaulay. They have been related to our present position by Lord Adrian, the present Master of Trinity College, Cambridge, in his 1961 lecture commemorating the 400th anniversary of Bacon's birth, entitled "Francis Bacon, the Advocate of Science." Lord Adrian reflected on Bacon's classic formulation of the scientific method, and then went on to say,

> .... but Bacon's writing on science had another and greater appeal to his own time, though it concerns an aspect which is now taken for granted. Macaulay and many who have followed him make this by far the most important feature of Bacon's system. He

J. B. FISK Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

insisted on the great practical value of scientific knowledge. The insight which the scientist obtains into nature can and should be employed in commanding nature for the service of man. Macaulay says that Bacon used means different from those of other philosophers because he wished to arrive at an end altogether different from theirs. The end was 'fruit' rather than 'light,' utility and progress in improving the condition of the human race, the good of mankind in the sense in which the mass of mankind has always understood the word 'good.' 'To make men perfect was no part of Bacon's plan. His humble aim was to make imperfect man comfortable.'

So now we see already a remarkable coordination in the mind and sayings of the philosopher Bacon, so generally accepted as the originator of the modern scientific method, with the idea of the application of science. Indeed, his method, embodying not only theoretical postulates (as from the Greeks), but also experimental trial and verification, is the very essence of modern pure science, and he boldly linked it with a motive of utility. Moreover, Bacon in his *Novum Organum* forwarded principles in this regard which would even outdo the modern project engineer (although our vantage point after nearly 300 years of scientific research and discovery would compel some specification of exact meaning). His statement was ". . . another powerful and great cause of the little advancement of the sciences which is this. It is impossible to advance properly in the course when the goal is not properly fixed. But the real and legitimate goal of the sciences is the endowment of human life with new inventions and riches." Of course, Bacon also valued highly what we have been calling the "light" in science as well as its "fruit" and Lord Adrian puts it all together well in his remark, "It is not cant to put truth above comfort, but most of us would like both if we could get them and Bacon's plan aimed at both, for he held them to be inseparable."

So we have joined historically the essence of the scientific method, whose triumphs we are celebrating once more in this Centennial, with the doctrine of the usefulness of science in human affairs. And our experience of the past century does support the view that these great elements of science — its method and the part of its meaning devoted to human use — are closely interconnected. So I shall try to bring out some examples from various domains of science and technology wherein the synthesis of scientific knowledge for human use seems to have reacted profoundly on *the quality of discovery* itself and especially on *the direction of inquiry* into various fruitful fields of research and invention.

Thus I should say at the very outset that, under suitable conditions, far from interfering with "science for its own sake," the applications of science seem steadily to be leading us into realms of greater and greater intellectual and even spiritual challenge. Similarly, as the complexity of science rises inexorably on its exponential scale (in terms of total information if not also in terms of basic principles), the vectors of applied science and technology do show directions in which pure scholars may couple to any degree they choose with the human issues and problems of their time. This, too, is not a bad thing for the motivation of men, or for smoothing the path between the ivory tower and public plaza.

My part is especially to discuss the *ways* in which this synthesis of scientific knowledge for use seems best to be achieved. At one stage, in our Academy's youth, it was said that professional engineering, coupled with a diligent search for invention, sufficed as ways to attain this objective. The successes of those methods are well known, especially in such cases as the vital development of the Bessemer process for steel making (including the concurrent patents of William Kelly of the Kelly-Pneumatic Process Company in this country!). The era of Edison was marked by this mode of applying new knowledge. Edison's conversations, as quoted by George Parsons Lathrop, develop this concept of the relations of discovery to invention. He said,[1]

> Discovery is not invention and I dislike to see the two words confounded. A discovery is more or less in the nature of an accident. . . . Goodyear discovered the way to make hard rubber. He was at work experimenting with india rubber, and quite by chance he hit upon a process which hardened it—the last result in the world that he wished or expected to attain. Bell's telephone was a discovery, too. He was engaged with the possibilities of sending sound waves over a telegraph wire, and found an invention by which this could be done. Then, by accident, it was discovered that articulate speech could be sent out over the wire— and there was the telephone. But Bell did not set out to make an instrument by which talk could be transmitted, and therefore I say he discovered instead of invented the telephone. In a discovery there must be an element of the accidental, and an important one too; while an invention is purely deductive. . . . In my own case but few, and those the least important, of my inventions owed anything to accident. Most of them have been hammered out after long and patient labor, and are the result of countless experiments, all directed toward attaining some well-defined object." Edison said then, concerning the very difficult work on the electric light, "And yet, through all those years of experimenting and research I never once made a discovery. All my work was deductive, and the results I achieved were those of invention pure and simple.

Thus we see that the specific goal-setting of engineering, which is of course vital to the magnificent successes of that profession, has long been differentiated from the method of discovery. Yet in the growing application of science for human use there have arisen increasingly satisfactory combinations of the subtle elements of scientific discovery, and of the hard-driving, problem-solving techniques of engineering. This is foreshadowed by Langmuir's remarks, in his essay on "fundamental research and its human values," concerning his discovery of the gas-filled electric light bulb. He said, "I want to show you how . . . the practical result could hardly have been reached in a laboratory in which the workers were assigned definite work. . . ." But Langmuir's research, carried out in a "practical" environment, led to a succession of triumphs.

What has been the key factor in merging these diverse attitudes, so that Bacon's original aspiration of science for the betterment of the human condition is being increasingly realized? I submit that, curiously enough, this element is frequently ignored in the highly personal evolution of the body of science itself. It is an institution of men, a structural community of experts. In the face of the universally accepted precept that scientific (and almost all other scholarly) achievement is the result of individual effort exerted by particular, gifted minds, it is striking that a community of effort constructed as an institution of men is the basis for the most spectacular progress we have made in the application of scientific knowledge for human use. You may at once indignantly reject this finding, and ask: How can you say that when the history of applied science is also that of individual genius — as in Pasteur's vaccines, in Bell's telephone, in Edison's light bulb, in Hall's aluminum process, and indeed in Chadwick's neutron? Truly these were the results of individual, personal insight and ingenuity. So I would quickly say that I am not talking about teams, as diligent and efficient as they have been in the application of knowledge in many vital areas such as synthesis of antimalarials and antibiotics, the evolution of microwave technology and radar, the engineering of nuclear energy, and so on. Rather I am saying that the best application of science and the synthesis of new knowledge turn out to be by a community of gifted people working intimately but independently, with each free to follow his own mind. Such a group constitutes the essential mechanism to bring together the diverse elements, many disciplines and kinds of learning from the most abstract mathematics to the most concrete mechanics and design.

Indeed it is interesting that, in the historic advance of science particularly in the past century, quite separate disciplines such as physics, chemistry, biology, and so forth were formed. Certain very distinct compartments were created even with those disciplines, such as organic chemistry, nuclear physics, molecular biology, etc. These divisions arose and represented domains in which by bold simplification and postulate scholars had been able to understand certain parts of nature. This naturally was also the way to teach this understanding, and so these classifications were retained and have become classic parts of the school and university structure.

But, of course, the needs of man at times are relatively unconnected with these particular categories of human knowledge. They are divisions of knowledge in which the most fruitful experiments and the most revealing theories could be exercised. For a long time this kind of classification extended also into engineering itself, as revealed by the designations: mechanical engineering, electrical engineering, and so on. More recently, terms like "industrial design" have become prevalent, and these begin to recognize the "human use" features of the profession. Medicine or medical science was probably the earliest recognition of such technical professions devoted to human needs, but itself was separated (and some say it still is) from the classical categories of biology, chemistry, and physics, as well as statistics and other mathematics, which provide its new and growing resources.

So this notion of combination of the sciences, or what is resonantly termed interdisciplinarianism nowadays, representing the free interaction of experts in a very wide range of scientific endeavor, is an institutional quality which has arisen almost entirely in the past few decades or at least half a century. It has also arisen very largely because of a new coupling of science with the human needs. Not surprisingly, such a structure is at last being found to be of great merit in approaching refractory problems of understanding nature for its own sake. There is a trend to have groups of many formally separate parts of science clustered around great nuclear accelerators or in genetics laboratories. And, of course, in an increasing variety of national scientific endeavors, such as space exploration, nuclear energy production, oceanography, and assaults against cancer, coronary disease, and other human afflictions, this combination of diverse specialities and its institutional cultivation are appearing.

Hence, it is useful on this occasion to recognize the relative newness of our experience with this kind of scientific and technical endeavor and to try to trace some of its origins and to suggest some of its projections. Strikingly, the philosopher Alfred North Whitehead, in his book *Science and the Modern World,* saw some of the essence of the matter in comments he made 40 years ago about the 19th century itself. He said,

> What is peculiar and new to the [19th] century, differentiating it from all its predecessors, is its technology. It was not merely the introduction of some great isolated inventions. It is impossible not to feel that something more than that was involved. . . . The process of change was slow, unconscious, and unexpected. In the nineteeth century, the process became quick, conscious, and expected. . . . The whole change has arisen from the new scientific information. Science, conceived not so much in its principles as in its results, is an obvious storehouse of ideas for utilisation. . . . Also, it is a great mistake to think that the bare scientific idea is the required invention, so that it has only to be picked up and used. An intense period of imaginative design lies between. One element in the new method is just the discovery of how to set about bridging the gap between the scientific ideas, and the ultimate product. It is a process of disciplined attack upon one difficulty after another. . . . This discipline of knowledge applies beyond technology to pure science, and beyond science to general scholarship. It represents the change from amateurs to professionals. . . . But the full self-conscious realisation of the power of professionalism in knowledge in all its departments, and of the way to produce the professionals, and of the importance of knowledge to the advance of technology, and of the methods by which abstract knowledge can be connected with technology, and of the boundless possibilities of technological advance,—the realisation of all these things was first completely attained in the nineteeth century. . . .

Now this attainment, referred to by Whitehead, sounds as though the scheme we are emphasizing as the principal way by which new knowledge is synthesized in science for human use was indeed old rather than relatively new. Actually, Whitehead had observed, specifically in Germany, certain highly developed forms of organizing to achieve particular technical ends, especially in the chemical and metallurgical industries. He shrewdly discerned the general outlines of the methods. I think he did not mean to say that there was much use of combinations of experts from various fields but rather that groups of experts of some sort were the basis of this revolutionary professional organization and use of new knowledge from science.

In our next considerations of how these communities of experts may be cultivated best to extend the synthesis of knowledge for human use, we should recall that James Clerk Maxwell, in delivering the Rede lecture nearly 80 years ago, chose to talk on The Telephone. Apparently what interested him was Bell's ability to bring together diverse sources of scientific and technical information, and even more to appreciate the points of view of various specialists in the relevant fields of electricity, acoustics, physiology, and others. Already, then, Maxwell saw the opportunities for combination of many new kinds of knowledge into useful results. Thus he said:[2]

> I shall, therefore, consider the telephone as a material symbol of the widely separated departments of human knowledge, the cultivation of which has led, by as many converging paths, to the invention of this instrument by Professor Graham Bell.
>
> For whatever may be said about the importance of aiming at depth rather than width in our studies, and however strong the demand of the present age may be for specialists, there will always be work, not only for those who build up particular sciences and write monographs on them, but for those who open up such communications between the different groups of builders as will facilitate a healthy interaction between them.

This was said 85 years ago.

Now we have found in the succeeding years in attempting, for example, to carry on the enterprise initiated by Bell's discovery of the telephone, that the cohesive element which provided this healthy interaction of Maxwell is a common and a tenable set of objectives. Curiously enough it is not a case where in the history of electrical communications; for instance, theories or important discoveries have themselves formulated the objectives of the enterprise. Rather, clear, simple, technical thinking about the human needs suggested goals of such value and challenge to the mind that communities of scientists and engineers were interested and willing to combine all sorts of skills and learnings in order to advance toward these goals. Thus the idea of nation-wide and continent-wide telephony, in which any person anywhere in the nation could through his own telephone speak to anyone else, was at the time it was put forward by Theodore Vail, a highly visionary statement. Yet it was a simple principle, meaningful technically, which has consistently and progressively energized the range of effort in our telephone laboratories, beginning with the exploitation of the qualities of electromagnetic induction found by Joseph Henry as embodied in the loading coil, and extending through to the creation of the vacuum tube by H. D. Arnold, the discovery of negative feedback by Harold Black, of the coaxial cable and carrier principles by Affel and Espenschied, advances in information theory by Claude Shannon, the invention of the transistor by Bardeen, Brattain, and Shockley, the discovery and exploitation of the traveling wave tube by Kompfner and Pierce, etc.

Along this route toward Vail's goal which, of course, has been somewhat elaborated into the notion of world-wide telephony and electrical communicating now, have arisen many byways. Many of them were tempting byways, such as the implementation of the first radio broadcasting, the realization of sound for moving pictures, the demonstration of full color video, even the discovery by one engineer working on telephone transmitters of the principles now embodied in the universally employed diaphragm fuel pump for automobile engines. An issue of importance for this gathering is that it was decided, each time with considerable agony, *not* to pursue these byways to the exclusion of the main pathway toward the central goal. There seems to be a good deal of power in this policy with respect to our central theme of cultivating communities of experts in science and technology so that they show optimum effectiveness in meeting the needs of man. There is the especially human tendency for experts to want to see developed their own special discoveries. Naturally, the chemists who found in the telephone laboratories a way to make laminating plastics out of previously unusable but highly stable and efficient polymers would like to see a generalized application in the form of materials and laminate manufacture with usual product development, etc. The physicists discovering electron diffraction in the telephone laboratories might even be glad to see a venture into instrument making of the widespread use of this technique. The engineer and acoustics expert finding ways to make sound tracks on moving pictures in the telephone laboratories would like to see a major venture into the mass entertainment–moving picture field result from this discovery. Innumerable examples could be cited and could be multiplied in many other laboratories.

Now the point is that as soon as these diverse paths of technical exploitation are opened up, and indeed they are often most attractive to managements in terms of quick profits and rapid invasion of a new market (which offhand seems importantly to reflect a new human need), nevertheless as soon as they are opened up, each of the community of experts begins to think: Now it is my turn to make something, some sort of discovery or invention or new level of understanding which will send our institution off into my particular direction. The great goals become fragmented and the original purpose of being able to turn a marvelous diversity of knowledge from all the independent and separate sciences and arts into a common, generally oriented effort toward major overriding goals for human uses, is badly obscured.

So I assert that the chief tenet for assembling scientific knowledge for human use is the formulation of objectives which are (*a*) sufficiently important, (*b*) suitably broad, and (*c*) technically meaningful so that they will provide a vast template of opportunities for the specialities and for the individualities of creative scientists and engineers, to inscribe the over-all effort without providing diversions. (The danger of these diversions, incidentally, may also apply strongly to our present national preoccupation with side benefits, or "fall-out," both economic and social, from major national exertions.)

Also the quality of being sufficiently important is one which technical goals for exploitation of scientific knowledge do not easily specify themselves. As other speakers in this session have shown there are also things which are only with great difficulty specified by the judgments of busy governments and bureaucracies. There is, however, particularly in our system of free enterprise, which has flourished so strongly in the century we are celebrating, a curiously pervasive scale of values. It is competition in the public market place. While the scientist and the engineer will never feel that his public understands all the values or rates properly the importance of his technology, nevertheless there is an increasing discernment by the public of the meaning of science and engineering for the betterment of man. There is a very strong indication that the kinds of human needs for whose satisfaction humanity is willing to pay give useful guides to major goals.

Thus, I have purposely touched on questions of economics, of politics, of statecraft, of philosophy which, of course, have been and will be treated by others.

But we come back always to a few basic precepts, which can lead us to evermounting advance in the synthesis and applications of scientific knowledge for human use. Paramount is this doctrine of the right objectives — important, broad, meaningful — which will provide both force and direction for progress in the uses of science. And according to the ideas we have forwarded, the generation of new knowledge, indeed the vigor of science itself, will gain from this process of seeking to apply it. For the great thing we have learned, especially in the past century of modern science, but even as it was forecast by Francis Bacon, is that scientific matters related to human needs can be and very often have been the most challenging and "interesting" questions of all science.

While the great upsurge in the study of genetics and molecular biology, indeed of the vital process itself as reflected in abnormalities, such as malignancy, are current examples, there are many earlier cases: Planck's interest in luminescence leading to the quantum theory; communications and radar efforts leading to microwave spectroscopy; the ceaseless quest for the sources of energy, which apparently underlay much of the basic curiosity about nuclear forces. Over and over we see that high intellectual as well as practical values have been provided where it was possible for gifted minds to follow the kinds of goals we have tried to characterize.

So it can be in the future that science and its companion, engineering, may make both noble and benign the lot of man on earth and in his reaching toward the heavens. But we must take thought to form the aims and to construct the institutions which extend what we have already learned about the synthesis and application of scientific knowledge.

There has been a feeling, especially in the past decade, that science is getting too big for man and that we would either be destroyed by malign effects produced by it or be confused and eventually suffocated by the sheer mass of data and infinity of concepts, which it would impose on our culture. The view with which we conclude here is a vastly happier outlook. It reveals, on the one hand, that only the full satisfaction of all human needs (a situation not immediately in prospect) should distract us from setting goals for human use of such quality as to absorb constructively and beneficially any and all the scientific findings to come.

On the other hand, we see the culture of science itself gaining from this endeavor, because the way of seeking these goals will demand an ever-increasing combination and even a fusion of scientific skills, from what have been called, up to now, "different" sciences. While the endless variety of approach to science must be zealously preserved, these differences between sciences will grow less and less. Many of us believe that this change itself will aid emergence of great unifying principles. (Think, for instance, of information theory, of the coding of electrical signals, of the coding of nerve impulses, of the coding of bases in nucleic acids and of their influence on protein configuration.) For those in science and engineering who seek to solve a worthy problem, and to reach an end which is important, broad, meaningful, are finding ways more and more to bring all kinds of minds and skills to work together.

Science in the Satisfaction of Human Aspiration

I. I. RABI

IN THIS CONFERENCE have participated some of the most eminent men of science in the land. Day after day the mysteries of life were laid bare, and antecedent to life the structure of matter and indeed of the universe were presented in dramatic and fascinating clarity. Although no attempt was made to make a real interdisciplinary connection between the various disciplines, nevertheless the juxtaposition of topics did a very great deal to show us the essential unity of the scientific disciplines however different their techniques. The first session on the History of the Universe went in progressive steps from the origin of the elements to the origins of life. In the session on the Nature of Matter we went from the almost philosophical consideration of the organization of the laws of nature to the organization of living matter. At this session, turning to less immediate questions, we have had a general overview of science from the aspects of communication and application both to industry and to public policy and welfare. We listened to great wisdom.

It is now my turn. The task assigned to me was to address myself to science in its more intimate relation to the individual: science and the satisfaction of human aspirations. The drives which cause men to become interested in science are almost as various as human personality. Science could not happen without a range of personalities and cultures. We see the interplay now of cultures in the successive contributions of Egypt, Babylon, Greece, Alexandria, Rome, Arabian-Italian Renaissance, and in the modern era Japan, America, and China, which should perhaps be all by itself. One of the main personality types is the collector and classifier who may interest himself in stamps, books, pictures, or, on the other hand, in phases of botany, geology, astronomy, or even in parts of physics such as spectroscopy. We should not use the term "mere" with respect to

I. I. RABI Department of Physics, Columbia University

these activities. High talents of intellect and insight can be devoted to these aspects of science. The hunt for a new species of plant or animal, a new spectral sequence, a new collection or system of galaxies can be both exciting and demanding. Ingenuity, persistence, and what may be called luck are necessary for success in these endeavors. The born hunter and alert observer will find success and satisfaction in this phase of science. We would be nowhere without this type of individual.

Basically this element of science satisfies more immediately the desire to discover and to know the facts of nature. Nature with its tremendous variety and charm captures the fancy and the spirit. To one person a fact is just a fact and of no further interest. Such a person may become a mathematician or a logician but he will never be a true poet or scientist. The scientist, the experimental scientist at least, shares with the poet and artist a feeling for the value for the immediate and the empirical face of nature. The geologist loves his bright and shiny stone, his curious fossils, just as the physicist can never cease to be charmed by a spectrogram or the delightful paradoxes of the motion of a spinning top.

This aspect of the pursuit of science satisfies a basic desire or aspiration just to know, to find out, or perhaps make order out of the otherwise chaotic jumble of immediate experience. It is an aspiration shared with all mankind but more with youth and childhood than with adults. In this sense scientists are just children who never grew up, who never lost the nagging urge to ask how, why, and what. Like children, who in all innocence and high excitement bring a dangerous spider into the house and frighten the wits out of their elders, the scientist emerges with a smallpox vaccine or an atomic bomb.

There is another facet of human aspiration which contributes to the various faces of science. This is man in his aspect of the maker or doer. Again this aspect is strongest in childhood but persists longer into adulthood since its immediate use is obvious. The use of tools, the arts and crafts, are the hallmarks by which we rank prehistoric and primitive civilizations. Of course, arts and crafts are not yet science, but share with science the manipulation of nature. The one is to satisfy the desire for material needs of food, shelter, decoration, and of course armament. This is the usual, the normal — the sort of thing which is immediately understood by men of maturity and judgment such as we find in Congress. The other, the less practical but in the end more powerful, is to manipulate nature not for immediate or material ends but for the purpose of providing new knowledge or the tools which could provide new knowledge. This is the method of science.

It is rare that this aspiration for discovery is sufficiently understood by any community to the degree of actually providing funds for this purpose. Either the curiosity of childhood is soon lost (perhaps this change may have an actual survival value for the race) or somehow those who possess the gift lack the ability to communicate the deep meaning, the excitement, and the satisfactions of scientific discovery. This question should be of the greatest concern to our scientific community. Unless the public shares in our aspirations and our satisfactions in the scientific enterprise, the pure scientific impulse will always have to be diluted and even distorted with immediate ends which in the last two decades have been primarily military.

I see nothing ignoble or degrading in the application of science to the defense of one's country, quite the contrary. On the other hand, when the support of science is tied both administratively and by public interest directly to military and other purely practical uses, one begins to feel that somehow we as scientists have failed to arouse either interest or understanding in the public mind. We have failed to satisfy the aspirations which they share with the scientist by taking them with us as we go further along the road of scientific accomplishment. This community of interest which we have with the youth of the country in considerable measure does not extend to the adult population. Surely science is not only for children young and old.

I will not infringe further on what I have hoped would be Dr. Oppenheimer's domain of science and communication, but go on to still another facet, perhaps the most significant of the scientific adventure, or scientific movement, or, as I prefer to call it, the scientific culture. In this, we share with the poet and artist the delight in immediate empirical experience with its aesthetic, emotional, and intellectual values but we go further, not to express this experience in the language of the heart, but in the prose of the catalogue, finding similarities in differences and presenting the results as an intellectual structure which can inspire pleasure and interest. We share with the practical world the manipulation of nature but not principally directed to useful ends or ends which are said to be useful. Our ends go further but we do share with the artisan and the engineer the pleasure of invention of novel combinations to achieve what had hitherto been difficult or impossible.

Our goal is a sort of bootstrap operation to utilize the tools of present knowledge to gain new knowledge, knowledge which we could hardly have foreseen or imagined. In these two illustrations I hoped to show that in our interaction with the world outside ourselves the scientific aspirations and satisfactions are basically aspirations which are shared by all. The scientists' satisfactions come in a special form which expresses itself in the desire to broaden and deepen our knowledge and understanding of all phenomena, but whereas the rest of mankind concentrates on man, his feelings and desires, the scientist tries to see the world as it really is or might be shorn of man's perhaps excessive preoccupation with himself. Clearly this is a quest which can never come to an end. Scientific curiosity will never be satisfied because it will never reach its goal to know all and understand all.

Such a goal and such an adventure will hardly satisfy the more prosaic and limited aspects of our human nature but it nevertheless has a nobility of a kind which in other fields has called forth some of the greatest manifestations of the human spirit.

The third aspect of science which I wish to explore with you, and which may be as I suggested the most significant, is one which we share with the humanities and with religion. Except for some periods of uneasy truce, science and religion have always been in conflict. Since the time of Galileo this conflict has sharpened. Many able men, both from the side of science and of religion, have assured us that there is no conflict between the two. On closer examination, it is apparent that the synthesis or bridge which they try to establish results only in a devaluation of both aspects of a powerful urge of the human spirit.

The urge to comprehend the visible and invisible universe and to find man's place within it is common to both science and religion. The conflict between science and religion is therefore more in the nature of a civil war between two parties with the same ultimate aims of comprehension and of submission to a higher order of knowledge and of insight.

In these matters religion has always taken the lead. Questions about man's place in the universe and his origins had to be given answers in each generation. The ancient Hebrews could not wait for the discovery of the neutron and the development of the theory of stellar evolution, or for Darwin, Morgan, Crick, and Watson to explain the variety of life and the origins of man and of the universe. The noble opening lines of Genesis cannot fail to move the most prosaic scientist even today.

By means of dramatic imagery and lofty poetic insight, religions have provided world systems more or less complete which gave immediate satisfaction to the yearnings of man for order in the world and guidance in his life. They gave a release from certain fears although they sometimes substituted others for those displaced. The great human quality of faith was always a basic prop to these religious structures. Religion and religious systems to be fully effective had to become established in law and custom and in a certain sense their statements had to become to be regarded as self-evident.

Compared to the eagle flight of religious thought, science is more like the humdrum earth-bound bulldozer. Where it has passed anyone can follow. Whereas religion is aristocratic and hierarchical, science is democratic and leveling. After the bulldozer has passed, many beautiful gardens and buildings enshrined in history and sentiment may be destroyed but the ground is ready for newer and perhaps even more beautiful interesting cultivation, or perhaps not. In any event, a newer generation gets a new start.

It has often been said that science gives man knowledge but does not tell him what to do with it. These prescriptions and values he is supposed to get from religion or from the so-called humanities. To borrow a phrase from a great scientist, Enrico Fermi, "This is not very true." The great writings of the humanist and holy religious writ can do much to incite men to noble and charitable action as well as to acts of folly and cruelty.

Science can make no such claim. What science seeks to do in its limited way is to delineate the results of action through psychology on the individual himself, through the behavioral sciences on the others, and, above all, to present a choice of means, leaving the decision to an informed act of "free will" in so far as the term still has a meaning. Only those who do not care to make their actions their own will say that science does not contribute to values.

The conflict between science and religion, between science and the humanities, therefore remains. The latter must always claim more than it knows and therefore must always retreat and qualify as science advances. "Don't insult me with facts" is the hurt expression of punctured pretensions. The true humanist and religionist welcomes scientific advance because it also allows him to advance his cause with deeper understanding. The Holy Fathers of the Roman Catholic Church have made this point very clear.

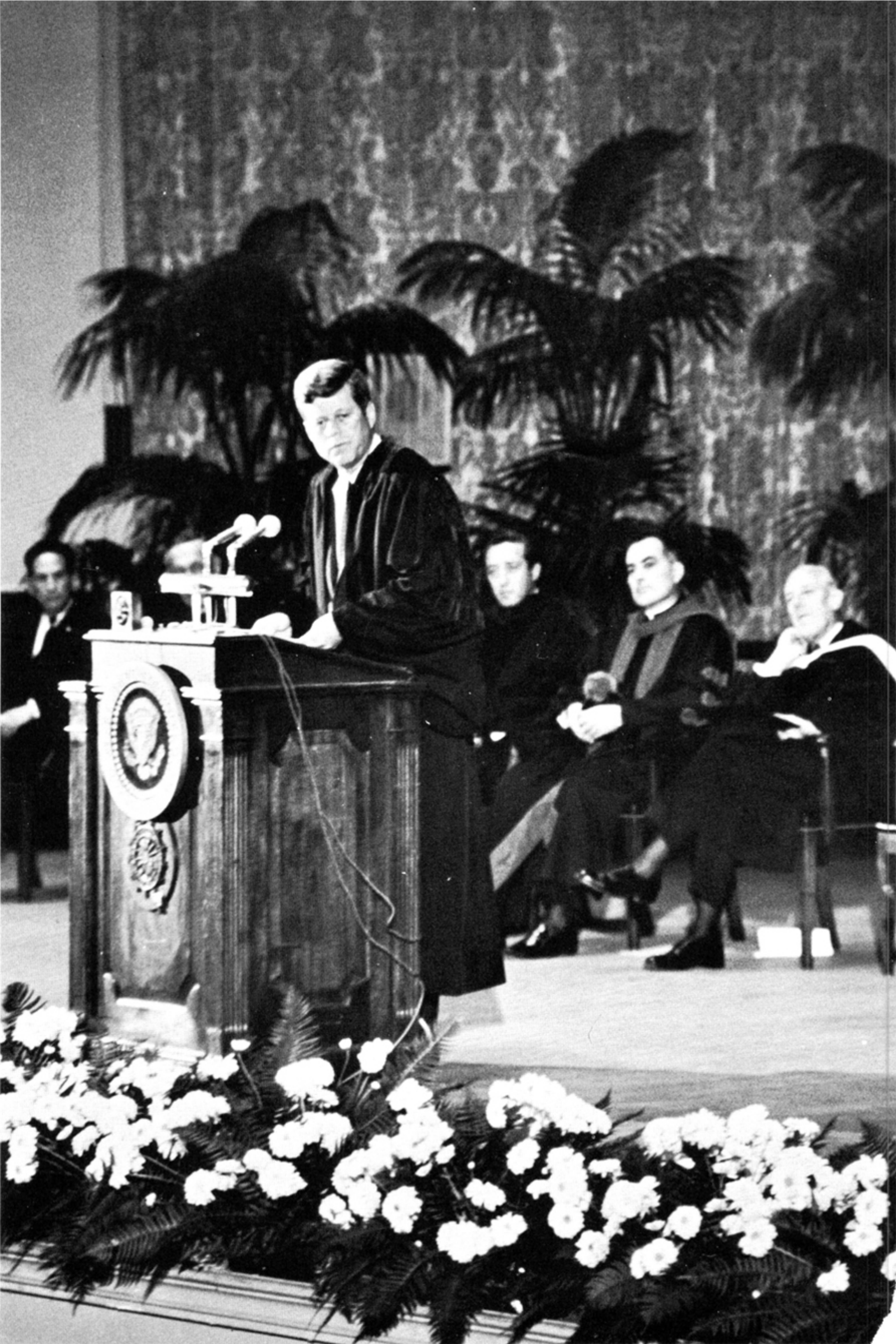
In contrasting science with the humanities and religion, the more pedestrian elements of the relentless march of scientific discovery have been emphasized. Actually science does not march in battle order toward a predetermined goal. This would be a contradiction in terms.

Scientists traditionally are free, untrammeled, and individualistic. Each sets his own goals following his interests. Such coordination as there is comes out of the nature of the subject matter and out of the tradition of the discipline; attempts to interfere, direct, or guide this freedom, as in some countries with overplanned societies, result in inefficiency and frustration of the creative urge. Scientists are well aware that they are prone to error. The observer in his laboratory knows full well that he can easily misinterpret his observation or miss the essential fact. The bold speculator can become so enamored of the beauty and sweep of his hypothesis that he may take it as an end in itself. It must be true, he feels, because it is so elegant. However, the court of highest appeal, which is nature itself, is relentless and error cannot long survive.

Therein lies one of the greatest appeals of science, an appeal which makes it capable of satisfying one of the greatest of human aspirations — to be a member of a community which is free but not anarchical. Science possesses an infinite variety of limited goals but in the end marches toward a limitless horizon. It consolidates its gains but does not rest on its laurels. Members of this community possess an inner solidity which comes from a sense of achievement and an inner conviction that the advance of science is important and worthy of their greatest effort. This solidity comes in a context of fierce competition, strongly held conviction, and differing assessments as to the value of one achievement or another. Over and above all this too human confusion is the assurance that with further study will come order and beauty and a deeper understanding.

One cannot close a discussion like this without bringing out one of the greatest rewards of the pursuit of scientific discovery. It comes accidentally and is often a matter of luck rather than the result of planning. It may come in an illuminating flash of insight or in the course of an experiment such as Rutherford's when he saw his alpha particles scattered through large angles or Anderson's when he saw an electron track moving the wrong way and realized he had a positive electron or Yukawa's when he saw that a supposed particle could account for nuclear forces. Although scientists don't write about these moments of exultation and ecstasy so different from the everyday routine of research, these fleeting visions can in one flash reward one for years of patient and exhausting work. At these times the scientist is filled with profound awe and humility that such wonders should be revealed through him. There is a quality about science, or rather about nature, which is always miraculous in its originality. To obtain a glimpse of this wonder can be the reward of a lifetime. This itself can be the sufficient satisfaction of the aspiration which makes scientists scientists.

At this point, at the end of my presentation and the end of this program, I can only wish the Academy and all mankind a century of achievement as great as the century which has passed in the life of our beloved Academy.



A Century of Scientific Conquest

President John F. Kennedy
addressing the Centennial Convocation of the
National Academy of Sciences
in Constitution Hall.
From the lectern, left to right,
Harrison Brown, The Reverend Theodore M. Hesburgh,
and Detlev W. Bronk.

A Century of Scientific Conquest

By JOHN F. KENNEDY

I AM HAPPY to accept the invitation to address the National Academy of Sciences, and I am very happy to come here with our distinguished visitor from Bolivia, the President of Bolivia, a distinguished scholar and educator in his own right, who in exile, has led his country through one of the most profound revolutions in the last decade that this hemisphere has witnessed. Therefore, I am proud that he is with me on this occasion very important to my own country. . . .

It is impressive to reflect that one hundred years ago, in the midst of a savage fraternal war, the United States Congress established a body devoted to the advancement of scientific research. The recognition then of the value of abstract science ran against the grain of our traditional preoccupation with technology and engineering.

You will remember de Tocqueville's famous chapter on why the Americans are more addicted to practical than to theoretical science. De Tocqueville concluded that, the more democratic a society, "the more will discoveries immediately applicable to productive industry confer on their authors gain, fame, and even power."

But if I were to name a single thing which points up the difference this century has made in the American attitude toward science, it would certainly be the wholehearted understanding today of the importance of pure science. We realize now that progress in technology depends on progress in theory; that the most abstract investigations can lead to the most concrete results, and that the vitality of a scientific community springs from its passion to answer science's most fundamental questions. I therefore greet this body with particular pleasure, for the range and depth of scientific achievement represented in this room constitutes the seedbed of our Nation's future.

The last hundred years have seen a second great change — the change in the relationship between science and public policy. To this new relationship, your own Academy has made a decisive contribution. For a century, the National

JOHN F. KENNEDY President of the United States of America

Academy of Sciences has exemplified the partnership between scientists who accept the responsibilities that accompany freedom, and a government which encourages the increase of knowledge for the welfare of mankind. As a result in large part of the recommendations of this Academy, the Federal Government enlarged its scientific activities through such agencies as the Geological Survey, the Weather Bureau, the National Bureau of Standards, the Forest Service, and many others, but it took the First World War to bring science into central contact with governmental policy and it took the Second World War to make scientific counsel an indispensable function of government. The relationship between science and public policy is bound to be complex.

As the country had reason to note in recent weeks during the debate on the test ban treaty, scientists do not always unite themselves on their recommendations to the makers of policy. This is only partly because of scientific disagreements. It is even more because the big issues so often go beyond the possibilities of exact scientific determination.

I know few significant questions of public policy which can safely be confided to computers. In the end, the hard decisions inescapably involve imponderables of intuition, prudence, and judgment.

In the last hundred years, science has thus emerged from a peripheral concern of government to being an active partner. The instrumentalities devised in recent times have given this partnership continuity and force. The question in all our minds today, is how science can best continue its service to the Nation, to the people, to the world, in the years to come.

I would suggest that science is already moving to enlarge its influence in three general ways: in the inter-disciplinary area, in the international area, and in the inter-cultural area. For science is the most powerful means we have for the unification of knowledge, and a main obligation of its future must be to deal with problems which cut across boundaries, whether boundaries between the sciences, boundaries between nations, or boundaries between man's scientific and his humane concern.

As science, of necessity, becomes more involved with itself, so also, of necessity, it becomes more international. I am impressed to know that of the 670 members of this Academy, 163 were born in other lands. The great scientific challenges transcend national frontiers and national prejudices. In a sense, this has always been true, for the language of science has always been universal and perhaps scientists have been the most international of all professions in their outlook, but the contemporary revolution in transport and communications has dramatically contributed to the internationalization of science, and one consequence has been the increase in organized international cooperation.

Every time you scientists make a major invention, we politicians have to invent a new institution to cope with it, and almost invariably these days and, happily, it must be an international institution. I am not just thinking of the fact that when you gentlemen figure out how to build a global satellite communications system, we have to figure out a global organization to manage it. I am thinking as well that scientific advantage provided the rationale for the World Health Organization and the Food and Agricultural Organization — that splitting the atom leads not only to a nuclear arms race, but to the establishment of the International Atomic Energy Agency; that the need for scientific exploration of Antarctica leads to an international treaty providing free access to the area without regard to territorial claims; that the scientific possibility of a World Weather Watch requires the attention of the World Meteorological Organization; that the exploration of oceans leads to the establishment of an Intergovernmental Oceanographic Commission.

Recent scientific advances have not only made international cooperation desirable, but they have made it essential. The ocean, the atmosphere, outer space, belong not to one nation or one ideology, but to all mankind, and as science carries out its tasks in the years ahead, it must enlist all its own disciplines, all nations prepared for the scientific quest, and all men capable of sympathizing with the scientific impulse.

Scientists alone can establish the objectives of their research, but society, in extending support to science, must take account of its own needs. As a layman, I can suggest only with diffidence what some of the major tasks might be on your scientific agenda, but I venture to mention certain areas which, from the viewpoint of the maker of policy, might deserve your special concern.

First, I would suggest the question of the conservation and development of our natural resources. In a recent speech to the General Assembly of the United Nations, I proposed a world-wide program to protect land and water, forests and wildlife, to combat exhaustion and erosion, to stop the contamination of water and air by industrial as well as nuclear pollution, and to provide for the steady renewal and expansion of the natural bases of life.

Malthus argued a century and a half ago that man, by using up all of his available resources, would forever press on the limits of subsistence, thus condemning humanity to an indefinite future of misery and poverty. We can now begin to hope and, I believe, know that Malthus was expressing not a law of nature, but merely the limitation then of scientific and social wisdom. The truth or falsity of his prediction will depend now, with the tools we have, on our own actions, now and in the years to come.

The earth can be an abundant mother to all of the people that will be born in the coming years if we learn to use her with skill and wisdom, to heal her wounds, replenish her vitality, and utilize her potentialities. And the necessity is now urgent and world-wide, for few nations embarked on the adventure of development have the resources to sustain an ever-growing population and a rising standard of living. The United Nations has designated this the Decade of Development. We all stand committed to make this agreeable hope a reality. This seems to me the greatest challenge to science in our times, to use the world's resources, to expand life and hope for the world's inhabitants. While these are essentially applied problems, they require guidance and support from basic science.

I solicit your help, and I particularly solicit your help in meeting a problem of universal concern — the supply of food to the multiplying mouths of a multiplying world. Abundance depends now on the application of sound biological analysis to the problems of agriculture. If all the knowledge that we now have were systematically applied to all the countries of the world, the world could greatly improve its performance in the low-yield areas, but this would not be enough, and the long-term answer to inadequate food production, which brings misery with it, must lie in new research and new experimentation, and the successful use of new knowledge will require close cooperation with other nations.

Already a beginning has been made. I think of the work in other countries, of the Rockefeller and Ford Foundations, and the creation by the OAS of the Inter-American Institute of Agricultural Sciences in Costa Rica. I look forward eventually to the establishment of a series of international agricultural research institutes on a regional basis throughout the developing world. I can imagine nothing more unwise than to hoard our knowledge and not disseminate it and develop the means of disseminating it throughout the globe.

Second, I would call your attention to a related problem; that is, the understanding and use of the resources of the sea. I recently sent to Congress a plan for a national attack on the oceans of the world, calling for the expenditure of more than $2 billion over the next ten years. This plan is the culmination of three years' effort by the Inter-Agency Committee on Oceanography, and it results from recommendations made by the National Academy.

Our goal is to investigate the world ocean, its boundaries, its properties, its processes. To a surprising extent, the sea has remained a mystery. Ten thousand fleets still sweep over it in vain. We know less of the oceans at our feet, where we came from, than we do of the sky above our heads. It is time to change this, to use to the full our powerful new instruments of oceanic exploration, to drive back the frontiers of the unknown in the waters which encircle our globe.

I can imagine no field among all those which are so exciting today than this great effort which our country and others will carry on in the years to come. We need this knowledge for its own sake. We want to know what is under the sea, and we need it to consider its bearings on our security, and on the world's social and economic needs. It has been estimated, for example, that the yield of food from the seas could be increased five or ten times through better knowledge of marine biology, and some day we will seed and weed and harvest the ocean. Here, again, the job can best be done by the nations of the world working together in international institutions.

As all men breathe the same air, so a storm along Cape Cod may well begin off the shores of Japan. The world ocean is also indivisible, and events in one part of the great sea have astonishing effects in remote places.

International scientific cooperation is indispensable if human knowledge of the ocean is to keep pace with human needs.

Third, there is the atmosphere itself, the atmosphere in which we live and breathe and which makes life on this planet possible. Scientists have studied the atmosphere for many decades, but its problems continue to defy us. The reasons for our limited progress are obvious. Weather cannot be easily reproduced and observed in the laboratory. It must, therefore, be studied in all of its violence wherever it has its way. Here, as in oceanography, new scientific tools have become available. With modern computers, rockets, and satellites, the time is ripe to harness a variety of disciplines for a concerted attack. And even more than oceanography, the atmospheric sciences require world-wide observation and, hence, international cooperation.

Some of our most successful international efforts have involved the study of the atmosphere. We all know that the World Meteorological Organization has been effective in this field. It is now developing a world-wide weather system to which nations the world over can make their contributions. Such cooperative undertakings can challenge the world's best efforts for decades to come.

Fourth, I would mention a problem which I know has greatly concerned many of you. That is our responsibility to control the effects of our own scientific experiments, for as science investigates the natural environment, it also modifies it, and that modification may have incalculable consequences for evil as well as for good.

In the past, the problem of conservation has been mainly the problem of human waste of natural resources, of their destruction, but science has the power for the first time in history now to undertake experiments with premeditation which can irreversibly alter our biological and physical environment on a global scale. The problem is difficult, because it is hard to know in advance whether the cumulative effects of a particular experiment will help or harm mankind. In the case of nuclear testing, the world is satisfied that radioactive contamination involves unnecessary risks, and we are all heartened that more than 100 nations have joined to outlaw testing in environments where the effects most directly threaten mankind.

In other fields we may be less sure. We must, for example, balance the gains of weather modification against the hazards of protracted drought or storm.

The Government has the clear responsibility to weigh the importance of large-scale experiments to the advance of knowledge or to national security against the possibility of adverse and destructive effects. The scientific community must assist the Government in arriving at rational judgments and interpreting these issues to the public. To deal with this problem, we have worked out formal procedures within the Government, to assure expert review before potentially risky experiments are undertaken, and we will make every effort to publish the data needed to permit open examination and discussion of proposed experiments by the scientific community before they are authorized.

If science is to press ahead in the four fields that I have mentioned, if it is to continue to grow in effectiveness and productivity, our society must provide scientific inquiry the necessary means of sustenance. We must, in short, support it. Military and space needs, for example, offer little justification for much work in what Joseph Henry called abstract science. Though such fundamental inquiry is essential to the future technological vitality of industry and Government alike, it is usually more difficult to comprehend than applied activity, and, as a consequence, often seems harder to justify to the Congress, to the Executive Branch, and to the people.

But if basic research is to be properly regarded, it must be better understood. I ask you to reflect on this problem and on the means by which, in the years to come, our society can assure continuing backing to fundamental research in the life sciences, the physical sciences, the social sciences, our natural resources, our agriculture, our protection against pollution and erosion. Together, the scientific community, the Government, industry, and education must work out the way to nourish American science in all its power and vitality. Even this year we have already seen in the first actions of the House of Representatives some failure of support for important areas of research which must depend on the national Government. I am hopeful that the Senate of the United States will restore these funds. Of course, what it needs is a wider understanding by the country as a whole of the value of this work which has been so sustained by so many of you.

I would not close, however, on a gloomy note, for ours is a century of scientific conquest and scientific triumph. If scientific discovery has not been an unalloyed blessing, if it has conferred on mankind the power not only to create, but also to annihilate, it has at the same time provided humanity with a supreme challenge and a supreme testing. If the challenge and the testing are too much for humanity, then we are all doomed, but I believe that the future can be bright, and I believe it can be certain. Man is still the master of his own fate, and I believe that the power of science and the responsibility of science have offered mankind a new opportunity not only for intellectual growth, but for moral discipline, not only for the acquisition of knowledge but for the strengthening of our nerve and our will.

We are bound to grope for a time as we grapple with problems without precedent in human history, but wisdom is the child of experience. In the years since man unlocked the power stored within the atom, the world has made progress, halting but effective, toward bringing that power under human control. The challenge, in short, may be our salvation. As we begin to master the potentialities of modern science, we move toward a new era in which science can fulfill its creative promise and help bring into existence the happiest society the world has ever known.

I express my appreciation to all of you for what you have done in your respective disciplines in the field of science, and for the contribution which those disciplines have made to the welfare of our country, and in the great sense, to the welfare of all mankind.

I can imagine no period in the long history of the world where it would be more exciting and rewarding than in the field today of scientific exploration. I recognize with each door that we unlock we see perhaps ten doors that we never knew existed and, therefore, we have to keep working forward, but with all of the tools now at our command, with all the areas of knowledge which are waiting to be opened up, I think that never in the short history of this Academy or in the far longer history of science has the time been brighter, the need been greater for the cooperation between those of us who work in Government and those of you who may work in far-distant laboratories on subjects almost wholly unrelated to the problems we now face in 1963. I hope that cooperation will remain intimate and that it will remain beneficial to both science and to the people as a whole.

Science has made all of our lives so much easier and happier in the last thirty years. I hope that the people of the United States will continue to sustain all of you in your work and make it possible for us to encourage other gifted young men and women to move into these high fields which require so much from them and which have so much to give to all of our people. So the need is very great. Even though some of your experiments may not bring fruition right away, I hope that they will be carried out immediately.

It reminds us of what the great French Marshal Lyautey once said to his gardener: "Plant a tree tomorrow." And the gardener said, "It won't bear fruit for a hundred years." "In that case," Lyautey said to the gardener, "plant it this afternoon." That is how I feel about your work.

References

HISTORY OF THE UNIVERSE [pages 3–134]

The Origin of the Elements

1. GAMOW, G., *Scientific American*, vol. 195:154 (1956).
2. SALPETER, E. E., *Tr. Internat. Astronom. Union*, vol. 10:661 (1960).
3. EDDINGTON, A., The Internal Constitution of the Stars, Cambridge, England, The University Press, 301 (1926).

Detailed bibliographies may be found in:

BURBIDGE, E. M., BURBIDGE, G. R., FOWLER, W. A., and HOYLE, F., *Rev. Mod. Physics*, vol. 29:547 (1957).

HOYLE, F., FOWLER, W. A., BURBIDGE, G. R., and BURBIDGE, E. M., *Astrophys. J.*, vol. 139:909 (1964).

The History of Stars and Galaxies

The history of the growth of concepts of stellar evolution, energy sources, and the modern picture of our Universe is too full for adequate reference here. A general background is given in Struve, O., and Zebergs, V., "Astronomy of the Twentieth Century," New York, The Macmillan Company, 1962.

The History of the Solar System

1. HENDERSON, G. H., and BATESON, S., *Proc. Roy. Soc. London, Series A*, vol. 145:563–591 (1934).
2. PATTERSON, C. C., Nuclear Processes in Geological Settings, Pennsylvania State University: *National Research Council, Pub. 400*, 157 (1955); *Geo. and Cosmochim. Acta*, vol. 10:230 (1956).
3. ANDERS, E., *in* The Moon, Meteorites and Comets, vol. 4 of The Solar System, (B. M. Middlehurst and G. P. Kuiper, editors), University of Chicago Press, 402 (1963).
4. FOWLER, W. A., The Origin of the Elements, *Proc. Nat. Acad. Sc.*, vol. 52:524–548 (1964).
5. OORT, J. H., *Bull. Astronomical Inst. Netherlands*, vol. 11:91 (1950).
6. WHIPPLE, F. L., *Astrophysic. J.*, vol. 111:375 (1950); vol. 113:464 (1951), *in* The Physics of Comets, Communications presented at the 4th International Colloquium on Astrophysics held at Liège, Sept. 19–21, 1952, *Mem. Soc. Roy. Sc., Liège*, vol. 13:321 (1953).
7. WHIPPLE, F. L., and JACCHIA, L. G., *Smithson. Contrib. Astrophys.*, vol. 4:97 (1961).
8. MCCROSKY, R. E., *Astronomical J.*, vol. 60:170 (abstract: 1955).
9. GOLDSTEIN, R. M., and CARPENTER, R. L., *Science*, vol. 139:910 (1963).
10. TER HAAR, D., and CAMERON, A. G. W., *in* Origin of the Solar System, (R. Jastrow and A. G. W. Cameron, editors), New York, Academic Press, Inc., 4 (1963).
11. SPITZER, LYMAN, JR., *Astrophysic. J.*, vol. 90:675 (1939).
12. VON WEIZSÄCKER, C. F., *Z. Astrophysik*, vol. 22: 319 (1944).
13. BIRKELAND, K., *Compt. rend. Acad. sc.*, vol. 155: 892 (1912).
14. BERLAGE, H. P., JR., *Proc. K. Akad. Amsterdam*, vol. 33:614, 719 (1930); vol. 35:553 (1932); vol. 37:221 (1934); vol. 38:857 (1935); vol. 43: 531, 557 (1940); vol. 51:796, 965 (1948).
15. ALFVÉN, H., *Stockholms Obs. Ann.*, vol. 14, No. 2, (1942); vol. 14, No. 5 (1943).
16. TER HAAR, D., *K. Danske Vidensk. Selsk. Mat-fys. Medd.*, vol. 25, No. 3 (1948).
17. MENZEL, D. H., The Universe in Action (The Rushton Lectures Foundation; Birmingham: Birmingham Printing Company [1957]); also private communication.
18. HOYLE, F., *Quart. J. Roy. Ast. Soc.*, vol. 1:28 (1960).
19. KUIPER, G. P., and MIDDLEHURST, B. M., editors, Planets and Satellites, vol. 3 of The Solar System, University of Chicago Press (1961).
20. WILDT, R., *Monthly Notices Roy. Astronomical Soc.*, vol. 107:84 (1947).
21. RAMSEY, W. H., *Monthly Notices Roy. Astronomical Soc.*, vol. 111:427 (1951); vol. 125:469 (1963).
22. DEMARCUS, W. C., *Astronomical J.*, vol. 63:2 (1958); and in *Handbuch der Physik*, (S. Flügge, editor), (Berlin, Springer-Verlag [1959]), vol. 52, Astrophysics III, The Solar System, p. 419.

23. UREY, H. C., The Planets, Their Origin and Development, New Haven, Yale University Press (1952).
24. UREY, H. C., *in* Physics and Astronomy of the Moon, (Z. Kopal, editor), New York, Academic Press, Inc., chapter 13 (1962).
25. BROWN, H., *Astrophysical J.*, vol. 111:641 (1950). BROWN, H., and INGRAHAM, M. G., *Physic. Rev.*, vol. 72:347 (1947).
26. GOLDBERG, L., MÜLLER, E. A., and ALLER, L. H., *Astrophysic. J.*, suppl., vol. 5, No. 45, 1 (1960).
27. SUESS, H. E., *in* Summer Course on Nuclear Geology, Varenna, 1960 (Comitato Nazionale per l'Energia Nucleare, Laboratorio di Geologia Nucleare, Pisa, Italy, 28 (1960)
28. KUIPER, G. P., *in* Astrophysics, (J. A. Hynek, editor), New York, McGraw-Hill Book Co., chapter 8 (1951).
29. RUSSELL, H. N., and MENZEL, D. H., *Proc. Nat. Acad. Sc.*, vol. 19:997 (1933).
30. RUBEY, W. W., *Bull. Geol. Soc. Am.*, vol. 62: 1111 (1951).
31. UREY, H. C., and CRAIG, H., *Geo. and Cosmochim. Acta*, vol. 4:36 (1953).
32. GOLES, G. G., FISH, R. A., and ANDERS, E., *Geo. and Cosmochim. Acta*, vol. 19:177 (1960).
33. NININGER, H. H., Out of the Sky, An Introduction to Meteoritics, New York, Dover Publications (1952). KRINOV, E. L., Principles of Meteoritics, New York, Pergamon Press (1960). MASON, B., Meteorites, New York, JohnWiley and Sons (1962).
34. SUESS, H. E., *Z. Elektrochem.*, vol. 53:237 (1949).
35. WOOD, J. A., JR., *Geo. and Cosmochim. Acta*, vol. 26:739 (1962); *Icarus*, vol. 2:152 (1963); and *in* The Moon, Meteorites and Comets, vol. 4 of The Solar System, (B. M. Middlehurst and G. P. Kuiper, editors), University of Chicago Press, chapter 12 (1963).
36. REYNOLDS, J. H., *Phys. Rev. Letters*, vol. 4:8 (1960).
37. FOWLER, W. A., GREENSTEIN, J. L., and HOYLE, F., *Geophys. J. Roy. Astronomical Soc.*, vol. 6:148 (1962).
38. MITLER, H., in press.
39. CAMERON, A. G. W., *Icarus*, vol. 1:13 (1962).
40. MESTEL, L., and SPITZER, L., *Monthly Notices Roy. Astronomical Soc.*, vol. 116:503 (1956).
41. SCHATZMAN, E., Special Tech. Rep. No. 3, Mt. Wilson and Palomar Observatories (1960).
42. UREY, H. C., *Proc. Chem. Soc.*, p. 67 (March, 1958).
43. LYTTLETON, R. A., *Monthly Notices Roy. Astronomical Soc.*, vol. 97:108 (1936).
44. MCCREA, W. H., *Proc. Roy. Soc. London, Series A*, vol. 260:152 (1961).
45. WHIPPLE, F. L., *Proc. Nat. Acad. Sc.*, in press.
46. JEFFREYS, SIR HAROLD, The Earth, Cambridge University Press, 4th edition (1959).
47. MOORE, H. J., private communication.
48. UREY, H. C., *in* Space Science, (D. P. LeGalley, editor), New York, John Wiley and Sons, chapter 4 (1963).
49. FREDRIKSSON, K., and RINGWOOD, A. E., *Geo. and Cosmochim. Acta*, vol. 27:639 (1963).
50. DUFRESNE, R. R., and ANDERS, E., *Geo. and Cosmochim. Acta*, vol. 26:1085 (1962).
51. LEVIN, B., L'Origine de la Terre et des Planètes, Moscow, Editions en langues étrangères (1958).
52. RINGWOOD, A. E., *Geo. and Cosmochim. Acta*, vol. 20:241 (1960).
53. WISE, D. U., *J. Geophysic. Research*, vol. 68:1547 (1963).
54. SUESS, H. E., *J. Geol.*, vol. 57:600 (1949).
55. MACDONALD, G. J. F., private communication.

I am grateful to Mrs. Jamie Diamandopoulos and Mr. Edward P. DeMatteo for art work on the figures, the former especially for Figure 1 and the latter especially for Figures 9, 13, and 14.

The Origins of Life

1. HOYLE, F., *Am. Scientist*, vol. 49:188 (1961).
2. HALDANE, J. B. S., *in* The Inequality of Man, Penguin Books (1937)
3. OPARIN, A. I., The Origin of Life, New York, Macmillan Co. (1938); 3rd edition, New York, Academic Press, Inc. (1957).
4. Recent general articles and books on this subject include: Origin of Life on the Earth, (A. I. Oparin, editor), London, Pergamon Press (1959). MOORE, P., and JACKSON, F., Life in the Universe, London, Routledge and Kegan Paul (1962). MILLER, S. L., and UREY, H. S., *Science*, vol. 130: 245 (1959). KEOSIAN, J., *Science*, vol. 131:479 (1960). LEDERBERG, J., *Science*, vol. 132:393 (1960).
5. HURLEY, P. M., How Old is the Earth? Garden City, New York, Anchor Books (1959).
6. TYLER, S. A., and BARGHOORN, E. S., *Science*, vol.

119:606 (1954). See also HARINGTON, T. S., and TOENS, P. D., *Nature,* vol. 200:947 (1963).

7. MILLER, S. L., *Science,* vol. 117:528 (1953); *J. Am. Chem. Soc.,* vol. 77:2351 (1955); *Biochim. et Biophysica Acta,* vol. 23:480 (1957).
8. ABELSON, P. H., *Science,* vol. 124:935 (1956).
9. HEYNS, K., WALTER, W., and MEYER, E., *Naturwissenschaften,* vol. 44:385 (1957).
10. ORÓ, J., *Biochem. Biophys. Res. Commun.,* vol. 2: 407 (1960). FOX, S. W., and HARADA, K., *Science,* vol. 133:1923 (1961).
11. FOX, S. W., and HARADA, K., *Science,* vol. 128: 1214 (1958). ORÓ, J., and GUIDRY, C. L., *Nature,* vol. 186:156 (1960).
12. UREY, H. C., *Proc. Nat. Acad. Sc.,* vol. 38:351 (1952); The Planets, New Haven, Yale University Press (1952).
13. MILLER, S. L., and UREY, H. C., *Science,* vol. 130: 245 (1959).
14. ABELSON, P. H., *Ann. New York Acad. Sc.,* vol. 69:276 (1957); also *in* Researches in Geochemistry, (P. H. Abelson, editor), New York, John Wiley and Sons, Inc. (1959).
15. RUBEY, W. W., *Geol. Soc. Am. Spec. Paper 62,* 631 (1955).
16. BASSHAM, J. A., and CALVIN, M., The Path of Carbon in Photosynthesis, Englewood Cliffs, New Jersey, Prentice-Hall, Inc. (1957). RACKER, E., *Nature,* vol. 175:249 (1955); *Harvey Lectures,* vol. 51:143 (1955–56).
17. ARNON, D. I., *Nature,* vol. 184:10 (1959).
18. RABINOWITSCH, E. I., Photosynthesis, New York, Interscience Press, Inc., vol. 1:7–10 (1945).
19. WALD, G., Abstracts of Papers, American Chemical Society Meeting in Boston, April, 1959, Division of Biological Chemistry, Communication 44, p. 22C.
20. HOROWITZ, N. H., *Proc. Nat. Acad. Sc.,* vol. 31: 153 (1945).
21. WALD, G., *Scient. Am.,* vol. 201:92 (October, 1959).
22. LOVELL, B., The Exploration of Outer Space, London, Oxford University Press, 70–72 (1962).
23. MACALLUM, A. B., *Physiol. Rev.,* vol. 6:316 (1926). For a quite different view of this situation, however, see CONWAY, E. J., *Proc. Roy. Irish Acad., Sect. B,* vol. 48:161 (1943).
24. BERG, G., Das Vorkommen der chemischen Elemente auf der Erde, Leipzig, J. A. Barth, 113 (1932).
25. WALD, G., *in* Horizons in Biochemistry, (M. Kasha and B. Pullman, editors), New York, Academic Press, Inc., 127 (1962).
26. HENDERSON, L. J., Fitness of the Environment, New York, Macmillan Co. (1913); reprint edition, Boston, Beacon Press (1958).
27. This experiment was first done for me by R. C. Bray and H. W. Dougherty, then graduate students in the Biochemistry Department at the Columbia University College of Physicians and Surgeons.
28. Small fractions of D-amino acids appear in a few strange places: several antibiotics and the capsular substance of the pathogenic bacterium *Bacillus anthracis.*
29. WALD, G., *Ann. New York Acad. Sc.,* vol. 69:352 (1957).
30. WALD, G., *in* Comparative Biochemistry, (M. Florkin and H. S. Mason, editors), New York, Academic Press, Inc., vol. 1:311 (1960).
31. WIGNER, E. P., *in* The Logic of Personal Knowledge, (Festschrift for Michael Polanyi), London, Routledge and Kegan Paul, chapter 19 (1961).

NATURE OF MATTER [pages 137–203]

Symmetry and Conservation Laws

1. G. HAMEL in his Theoretische Mechanik (B. G. Teubner, 1912) mentions (page 130) Jordanus de Nemore ($\sim$ 1300) as having recognized essential features of what we now call mechanical energy and Leonardo da Vinci as having postulated the impossibility of the perpetuum mobile.
2. F. CAJORI's History of Physics (New York, Macmillan Company, 1929) gives exactly half a line to it (page 108).
3. See, for instance, his semipopular booklet Relativitätstheorie (Braunschweig, Friedrich Vieweg und Sohn, various editions, 1916–1956).
4. HAMEL, G., *Z. Math. Phys.,* vol. 50:1 (1904); also ENGEL, F., *Abhandl. Ges. Wissensch. Göttingen, Math.-Physik. Kl.,* 270 (1916).
5. See the present writer's article in *Progr. Theoret.*

Phys., vol. 11:437 (1954); also MURAI, Y., *Progr. Theoret. Phys.*, vol. 11:441 (1954); and more recently, GREENBERGER, D. M., *Ann. Phys.*, vol. 25: 290 (1963).

6. UTIYAMA, R., *Phys. Rev.*, vol. 101:1597 (1956); also YANG, C. N., and MILLS, R. L., *Phys. Rev.*, vol. 96:191 (1954).
7. POINCARÉ, H., *Compt. rend. Soc. biol.*, vol. 140: 1504 (1905); *Rend. Circ. Math. Palermo*, vol. 21: 129 (1906).
8. Thus Aristotle's Physics postulated that motion necessarily required the continued operation of a cause. Hence, all bodies would come to an absolute rest if they were removed from the cause which imparts them a velocity (cf. e.g. CROMBIE, A. C., Augustine to Galileo, London, Falcon Press, 1952, page 82 or 244). This cannot be true for coordinate systems moving with respect to each other. The coordinate systems with respect to which it is true then have a preferred state of motion.
9. MACH, E., The Science of Mechanics (various editions, Chicago, Open Court Publishing Company), chapter III, section 3.
10. FOCK, V., The Theory of Space, Time and Gravitation, New York, Pergamon Press (1959). Similar ideas were proposed before by KRETSCHMAN, E., *Ann. Physik*, vol. 53:575 (1917). See also EINSTEIN, A., *Ann. Physik*, vol. 55:241 (1918).
11. For the strong interaction, compare NE'EMAN, Y., *Nucl. Phys.*, vol. 26:222 (1961), and GELL-MANN, M., *Phys. Rev.*, vol. 125:1067 (1962). For the weak interaction, FEYNMAN, R. P., and GELL-MANN, M., *Phys. Rev.*, vol. 109:193 (1958), and SUDERSHAN, E. C. G., and MARSHAK, R. E., *Phys. Rev.*, vol. 109:1960 (1958); also SAKURAI, J. J., *Nuovo Cimento*, vol. 7:649 (1958), and GERSHTEIN, G. S., and ZELDOVITCH, A. B., *J. Exp. Theor. Phys. USSR*, vol. 29:698 (1955).
12. For the baryon conservation law and the strong interaction, this was suggested by the present writer, *Proc. Am. Phil. Soc.*, vol. 93:521 (1949), and *Proc. Nat. Acad. Sc.*, vol. 38:449 (1952). The baryon conservation law was first postulated by STUECKELBERG, E. C. G., *Helv. Phys. Acta*, vol. 11:299 (1938).
13. For the experimental verification of these and the other conservation laws, see FEINBERG, G., and GOLDHABER, M., *Proc. Nat. Acad. Sc.*, vol. 45: 1301 (1959). The conservation law for leptons was proposed by MARX, G., in *Acta Phys. Hung.*, 3:55 (1953); also ZELDOVITCH, A. B., *Dok. Akad. Nauk SSSR*, vol. 91:1317 (1953), and KONOPINSKI, L. J., and MAHMOUD, H. M., *Phys. Rev.*, vol. 92:1045 (1953). It seemed to be definitely established by LEE, T. D., and YANG, C. N., *Phys. Rev.*, vol. 105:1671 (1957).
14. For the baryon conservation and strong interaction, this was emphatically pointed out in a very interesting article by SAKURAI, J. J., *Ann. Phys.*, vol. 11:1 (1960). Concerning the conservation of lepton member, see MARX, G., *Z. Naturforsch.*, vol. 9a:1051 (1954).
15. GOLDBERGER, M. L., *Phys. Rev.*, vol. 99:979 (1955); GELL-MANN, M., and GOLDBERGER, M. L., *Phys. Rev.*, vol. 96:1433 (1954). See also GOLDBERGER, M. L., and WATSON, K. M., Collision Theory, New York, John Wiley and Son, chapter 10 (1964).

The Architecture of Molecules

1. BARLOW, W., *Nature*, vol. 29:186, 205, 404 (1883); *Z. Krist.*, vol. 23:1 (1894); vol. 29:433 (1898).
2. SLATER, J. C., *Phys. Rev.*, vol. 37:481 (1931).
3. KEMP, J. D., and PITZER, K. S., *J. Am. Chem. Soc.*, vol. 59:276 (1937).
4. KASPER, J. S., LUCHT, C. M., and HARKER, D., *Acta Cryst.*, vol. 3:436 (1950).
5. WUNDERLICH, J. A., and LIPSCOMB, W. N., *J. Am. Chem. Soc.*, vol. 82:4427 (1960).
6. SAMSON, S., *Nature*, vol. 195:259 (1962).
7. LATIMER, W. M., and RODEBUSH, W. H., *J. Am. Chem. Soc.*, vol. 42:1419 (1920).
8. COREY, R. B., *J. Am. Chem. Soc.*, vol. 60:1598 (1938).
9. WATSON, J. D., and CRICK, F. H., *Nature*, vol. 171:737, 964 (1953).
10. PAULING, L., and COREY, R. B., *Arch. Biochem. Biophys.*, vol. 65:164 (1956).

The Organization of Living Matter

1. Mostly C, N, H, O, P, and S supplemented with a series of trace elements.
2. In the morphological sense.
3. The abbreviations used in this article are: DNA, deoxyribonucleic acid; RNA, ribonucleic acid; ATP, adenosinetriphosphate.
4. MICHAELIS, L., *Arch. f. mikroskop. Anat. u. Entwicklungsmechn.*, vol. 55:558 (1900).

5. Palade, G. E., *J. Histochem. and Cytochem.,* vol. 1:188 (1953). Sjöstrand, F. S., and Rhodin, J., *Exp. Cell Research,* vol. 6:426 (1953).
6. Fernández-Morán, H., *Circulation,* vol. 26:1039 (1962).
7. Claude, A., *J. Exp. Med.,* vol. 84:51, 61 (1946). Hogeboom, G. H., Schneider, W. C., and Palade, G. E., *J. Biol. Chem.,* vol. 172:619 (1948).
8. Hogeboom, G. H., Claude, A., and Hotchkiss, R. D., *J. Biol. Chem.,* vol. 165:615 (1946).
9. Kennedy, E. P., and Lehninger, A. L., *J. Biol. Chem.,* vol. 179:957 (1949).
10. Schneider, W. C., *J. Biol. Chem.,* vol. 176:259 (1948). Schneider, W. C., and Potter, V. R., *J. Biol. Chem.,* vol. 177:893 (1949).
11. In retrospect, it was realized that the characteristic staining of mitochondria with Janus green was due to the local concentration of oxidative enzymes, *cf.* Lazarrow, A., and Cooperstein, S. J., *J. Histochem. and Cytochem.,* vol. 1:234 (1953).
12. Lehninger, A. L., *Harvey Lectures,* vol. 49:176 (1953–54). Green, D. E., *Harvey Lectures,* vol. 52: 177 (1956–57). Chance, B., and Williams, G. R., *Advances Enzymol.,* vol. 17:65 (1956).
13. Estabrook, R. W., and Holowinsky, A., *J. Biophysic. and Biochem. Cytol.,* vol. 9:19 (1961).
14. Blair, P. V., Oda, T., Green, D. E., and Fernández-Morán, H., *Biochemistry,* vol. 2:756 (1963).
15. Chance, B., Parsons, D. F., and Williams, G. R., *Science,* vol. 143:136 (1964).
16. Brierley, G. P., Bachmann, E., and Green, D. E., *Proc. Nat. Acad. Sc.,* vol. 48:1928 (1962). De Luca, H. F., Engstrom, G. W., and Rasmussen, H., *Proc. Nat. Acad. Sc.,* vol. 48:1604 (1962). Vasington, F. D., and Murphy, J. V., *J. Biol. Chem.,* vol. 237:2670 (1962).
17. Parsons, D. F., *Science,* vol. 140:985 (1963). Stoeckenius, W., *J. Cell Biol.,* vol. 17:443 (1963).
18. Lehninger, A. L., *J. Biol. Chem.,* vol. 190:345 (1951). Van den Bergh, S. G., and Slater, E. C., *Biochem. J.,* vol. 82:362 (1962).
19. Lehninger, A. L., Rossi, C. S., and Greenawalt, J. W., *Biochem. and Biophysic. Research Commun.,* vol. 10:444 (1963).
20. Peachey, L. D., *J. Cell Biol.,* vol. 20:95 (1964).
21. Garnier, Ch., *J. anat. et physiol.,* vol. 36:22 (1900).
22. Mathews, A., *J. Morphol.,* vol. 15, suppl., 171 (1899).
23. Porter, K. R., *J. Exp. Med.,* vol. 97:727 (1953); and The ground substance; observations from electron microscopy, *in* The Cell, (J. Brachet and A. E. Mirsky, editors), New York, Academic Press, Inc., vol. 2:621 (1961). Palade, G. E., *J. Biophysic. and Biochem. Cytol.,* vol. 2, No. 4, suppl., 85 (1956).
24. Palade, G. E., *J. Biophysic. and Biochem. Cytol.,* vol. 1:59 (1955). Palade, G. E., and Siekevitz, P., *J. Biophysic. and Biochem. Cytol.,* vol. 2:171, 671 (1956).
25. Dalton, A. J., and Felix, M. D., *Am. J. Anat.,* vol. 94:171 (1954). Dalton, A. J., Golgi apparatus and secretion granules, *in* The Cell, (J. Brachet and A. E. Mirsky, editors), New York, Academic Press, Inc., vol. 2, p. 603 (1961).
26. Watson, M. L., *J. Biophysic. and Biochem. Cytol.,* vol. 1:257 (1955).
27. Caro, L. G., and Palade, G. E., *J. Cell Biol.,* vol. 20:473 (1964).
28. de Duve, Ch., Pressman, B. C., Gianetto, R., Wattiaux, R., and Appelmans, F., *Biochem. J.,* vol. 60:604 (1955). de Duve, Ch., The lysosome concept, *in* Ciba Foundation Symposium on Lysosomes, (A. V. S. de Reuck and M. P. Cameron, editors), London, J. & A. Churchill Ltd., p. 1 (1963). Novikoff, A. B., Lysosomes in the physiology and pathology of cells; contributions of staining methods, *in* Ciba Foundation Symposium on Lysosomes, (A. V. S. de Reuck and M. P. Cameron, editors), London, J. & A. Churchill, Ltd., p. 36 (1963). Novikoff, A. B., Lysosomes and related particles, *in* The Cell, (J. Brachet and A. E. Mirsky, editors), New York, Academic Press, Inc., vol. 2:423 (1961).
29. Palay, S. L., and Palade, G. E., *J. Biophysic. and Biochem. Cytol.,* vol. 1:69 (1955).
30. Warner, J. R., Rich, A., and Hall, C. E., *Science,* vol. 138:1399 (1962). Wettstein, F. O., Straehelin, T., and Noll, H., *Nature,* vol. 197: 430 (1963).
31. Roberts, B. R., editor, Microsomal Particles and Protein Synthesis, New York, Pergamon Press (1958). Ts'o, P. O. P., *Ann. Rev. Plant Physiol.,* vol. 13:45 (1962).
32. Jacob, F., and Monod, J., *J. Molecular Biol.,* vol. 3:318 (1961).
33. Hoagland, M., *Biochim. et Biophysica Acta,* vol. 16:288 (1955). Hoagland, M., Keller, E. B.,

and ZAMECNIK, P. C., *J. Biol. Chem.*, vol. 218:345 (1956).

34. HOLLEY, R. W., *J. Am. Chem. Soc.*, vol. 79:658 (1957). HOAGLAND, M. B., STEPHENSON, M. L., SCOTT, J. F., HECHT, L. I., and ZAMECNIK, P. C., *J. Biol. Chem.*, vol. 231:241 (1958). ZAMECNIK, P. C., *Harvey Lectures*, vol. 54:256 (1958–59).
35. NATHANS, D., ALLENDE, J. E., CONWAY, T. W., SPYRIDES, G. J., and LIPMANN, F., Symposium on Informational Macromolecules, New York, Academic Press, Inc., p. 349 (1963). KELLER, E. B., and FERGER, M. F., *Fed. Proc.*, vol. 23:164 (1964).
36. WEBSTER, G., and LINGREL, J. B., Protein synthesis by isolated ribosomes, *in* Protein Biosynthesis, (R. J. C. Harris, editor), London and New York, Academic Press, Inc., p. 301 (1961). RAAKE, J. D., *Biochim. et Biophysica Acta*, vol. 34:1 (1959).
37. LE PAGE, G. A., and SCHNEIDER, W. C., *J. Biol. Chem.*, vol. 176:1021 (1948). JOHNSON, M. R., *Biochem. J.*, vol. 77:610 (1960).
38. NEWBERG, R. W., and CHELDELIN, V. H., *J. Biol. Chem.*, vol. 218:89 (1956).
39. PORTER, K. R., and BRUNI, C., *Cancer Research*, vol. 19:997 (1959); *Anat. Rec.*, vol. 136:260 (1960). REVEL, J. P., NAPOLITANO, L., and FAWCETT, D. W., *J. Biophysic. and Biochem. Cytol.*, vol. 8:575 (1960).
40. Among subcellular components not covered by this discussion are: microbodies, multivesicular bodies, spindle fibers (tubules), centrioles, cilia, basal bodies, and microtubules, to name only the most frequently encountered.
41. MIESCHER, F., *Hoppe-Seyler med. chem. Untersuch.*, vol. 4:441 (1871).
42. ROODYN, D. B., *Internat. Rev. Cytol.*, vol. 8:279 (1959). ALLFREY, V. The isolation of subcellular components, *in* The Cell, (J. Brachet and A. E. Mirsky, editors), New York, Academic Press, Inc., vol. 1:193 (1959).
43. MAGGIO, R., SIEKEVITZ, P., and PALADE, G. E., *J. Cell Biol.*, vol. 18:267, 293 (1963).
44. KEIR, H. M., SMELLIE, R. M. S., and SIEBERT, G., *Nature*, vol. 196:752 (1962). SMITH, M. J., and KEIR, H. M., *Biochim. et Biophysica Acta*, vol. 68: 578 (1963).
45. WEISS, S. B., and GLADSTONE, L., *J. Am. Chem. Soc.*, vol. 81:4118 (1959). WEISS, S. B., *Proc. Nat. Acad. Sc.*, vol. 46:1020 (1960).
46. PRESCOTT, D. M., RNA synthesis in the nucleus and RNA transfer to the cytoplasm in *Tetrahymena piriformis*, *in* Biological Structure and Function, (T. W. Goodwin and O. Lindberg, editors); New York, Academic Press, Inc., vol. 2:527 (1961). *Cf.* also YANKOFSKY, S. A., and SPIEGELMAN, S., *Proc. Nat. Acad. Sc.*, vol. 49:538 (1963).
47. HUANG, R. C., and BONNER, J., *Proc. Nat. Acad. Sc.*, vol. 48:1216 (1962). ALLFREY, V. G., and MIRSKY, A. E., *Proc. Nat. Acad. Sc.*, vol. 48:1591 (1962). FRENSTER, J. H., ALLFREY, V. G., and MIRSKY, A. E., *Proc. Nat. Acad. Sc.*, vol. 50:1026 (1963).
48. ALLFREY, V. G., and MIRSKY, A. E., Amino acid transport into the cell nucleus and reactions governing nuclear protein synthesis, *in* Protein Biosynthesis, (R. J. C. Harris, editor), London and New York, Academic Press, Inc., p. 49 (1961).
49. STEPHENSON, M. L., THIMANN, K. V., and ZAMECNIK, P. L., *Arch. Biochem. and Biophysics*, vol. 65:194 (1956). APP, A. A., and JAGENDORF, A. T., *Biochim. et Biophysica Acta*, vol. 76:286 (1963). EISENSTADT, J., and BRAWERMAN, G., *Biochim. et Biophysica Acta*, vol. 76:319 (1963).
50. ROODYN, D. B., REIS, P. J., and WORK, T. S., *Biochem. J.*, vol. 80:9 (1961). ROODYN, D. B., SUTTIE, J. W., and WORK, T. S., *Biochem. J.*, vol. 83:29 (1962). KROON, A. M., *Biochim. et Biophysica Acta*, vol. 72:391 (1963).
51. CHUN, E. H. L., VAUGHAN, M. H., JR., and RICH, A., *J. Molecular Biol.*, vol. 7:130 (1963). SAGER, R., and ISHIDA, M. R., *Proc. Nat. Acad. Sc.*, vol. 50:725 (1963). KIRK, J. T. O., *Biochim. et Biophysica Acta*, vol. 76:417 (1963). GIBOR, A., and IZAWA, M., *Proc. Nat. Acad. Sc.*, vol. 50:1164 (1963). NASS, M., and NASS, S., *J. Cell. Biol.*, vol. 19:593, 613 (1963).
52. RIS, H., and PLAUT, E., *J. Cell Biol.*, vol. 13:383 (1962).
53. ROBERTSON, J. D., *Progr. Biophys. Biophysic. Chem.*, vol. 10:343 (1960).
54. WILBRANDT, W., *Internat. Rev. Cytol.*, vol. 13: 203 (1962). CHRISTENSEN, H. N., Biological Transport, New York, W. A. Benjamin, Inc. (1962).
55. SJÖSTRAND, F. S., *J. Ultrastructure Research*, vol. 8:517 (1963). FARQUHAR, M. G., and PALADE, G. E., *J. Cell Biol.*, vol. 17:375 (1963).
56. ESSNER, E., NOVIKOFF, A. B., and MASEK, B., *J.*

Biophysic. and Biochem. Cytol., vol. 4:711 (1958). FARQUHAR, M. G., and PALADE, G. E., *Proc. Nat. Acad. Sc.*, vol. 51:569 (1964).

57. ROBERTSON, J. D., *Biochem. Soc. Symp.*, vol. 16:3 (1959). YAMAMOTO, T., *J. Cell Biol.*, vol. 17:413 (1963).
58. ERNSTER, L., SIEKEVITZ, P., and PALADE, G. E., *J. Cell Biol.*, vol. 15:541 (1962). NOVIKOFF, A. B., ESSNER, E., GOLDFISCHER, S., and HEUS, M., Nucleosidephosphatase activities in cytomembranes, *in* The Interpretation of Ultrastructure, (R. J. C. Harris, editor), New York & London, Academic Press, Inc., p. 149 (1962).
59. TEDESCHI, H., and HARRIS, D. L., *Biochim. et Biophysica Acta*, vol. 28:392 (1958). TEDESCHI, H., JAMES, J., and ANTHONY, W., *J. Cell Biol.*, vol. 18:503 (1963).
60. SAGER, R., and PALADE, G. E., *J. Biophysic. and Biochem. Cytol.*, vol. 3:463 (1957).
61. GRANICK, S., The chloroplast: inheritance, structure and function, *in* The Cell, (J. Brachet and A. E. Mirsky, editors), New York, Academic Press, Inc., vol. 2:489 (1961).
62. LUCK, D. J. L., *J. Cell Biol.*, vol. 16:483 (1963). ZALOKAR, M., *J. Biophysic. and Biochem. Cytol.*, vol. 9:731 (1961). BLONDEL, B., and TURIAN, G., *J. Biophysic. and Biochem. Cytol.*, vol. 7:207 (1960).
63. ROBINOW, C. F., Outline of the visible organization of bacteria, *in* The Cell, (J. Brachet and A. E. Mirsky, editors), New York, Academic Press, Inc., vol. 4:45 (1961). RYTER, A., and KELLENBERGER, E., *Z. Naturforsch.*, vol. 13b:597 (1958).
64. A similar structure in a cell unit of even small dimensions (d $\sim$ 0.2μ) has been reported in Mycoplasmataceae (microorganisms of pleuropneumonia group) which at present are assumed to be the smallest free-living cells (VAN ITERSON, W., *J. Ultrastructure Research*, vol. 3:282 [1960]).
65. GLAUERT, A. M., and HOPWOOD, D. A., *J. Biophysic. and Biochem. Cytol.*, vol. 7:479 (1960). VAN ITERSON, W., *J. Biophysic. and Biochem. Cytol.*, vol. 9:183 (1961). FITZ-JAMES, P. C., *J. Biophysic. and Biochem. Cytol.*, vol. 8:507 (1960).
66. VANDER WINKEL, E., and MURRAY, R. G. E., *J. Ultrastructure Research*, vol. 7:185 (1962). VAN ITERSON, W., *J. Cell Biol.*, vol. 20:361 (1964).
67. WEIBULL, C., *J. Bact.*, vol. 66:696 (1953). MARR, A. G., Localization of enzymes in bacteria, *in* The Bacteria, (J. C. Gunsalus and R. Stanier, editors), New York, Academic Press, Inc., vol. 1:443 (1960).
68. PARK, R. B., and PON, N. G., *J. Molecular Biol.*, vol. 3:1 (1961).
69. CHANCE, B., ESTABROOK, R. W., and LEE, C. P., *Science*, vol. 140:379 (1963).

Genetic Determinants

1. BEADLE, G. W., Genetics and modern biology, Jayne Lectures for 1962, *Mem. Am. Phil. Soc.*, vol. 57 (October, 1963).
2. DOTY, P., *Harvey Lectures*, vol. 55:103 (1959–60).
3. TAYLOR, J. H., *Am. Scientist*, vol. 48:365 (1960).
4. REICH, E., *Cancer Research*, vol. 23:1428 (1963).
5. NIRENBERG, M. W., and MATTHAEI, J. H., *Proc. Nat. Acad. Sc.*, vol. 47:1588 (1961).
6. MARGOLIASH, E., *Proc. Nat. Acad. Sc.*, vol. 50:672 (1963).
7. INGRAM, V. M., Hemoglobins in Genetics and Evolution, New York, Columbia University Press (1963).
8. YANOFSKY, C., *in* Informational Macromolecules, New York, Academic Press, Inc., 195 (1963).
9. NIRENBERG, M. W., and JONES, O. W., JR., *in* Informational Macromolecules, New York, Academic Press, Inc., 451 (1963).
10. BENZER, S., *Harvey Lectures*, vol. 56:1 (1960–61).
11. FINCHAM, J. R. S., and DAY, P. R., *in* Fungal Genetics, Botanical Monographs, vol. 4 (Davis, 1963).
12. JACOB, F., and MONOD, J., *in* Cellular Regulatory Mechanisms, *Cold Spring Harbor Symp. Quant. Biol.*, vol. 26:193 (1961).
13. SAGER, R., and RYAN, F. J., Cell Heredity, New York, John Wiley and Sons, Inc. (1961).
14. JACOB, F., and WOLLMAN, E. L., Sexuality and the Genetics of Bacteria, New York, Academic Press, Inc. (1961).

THE DETERMINANTS AND EVOLUTION OF LIFE [pages 205–265]

The Differentiation of Cells

This is Contribution No. 740 from the Department of Zoology, Indiana University. The work of the author and his colleagues Janine Beisson-Schecroun, Laura Bukovsan, Ruth Dippell, and Ian Gibson was supported by grants from The Rockefeller Foundation and the American Cancer Society (Grant No. E 80A) and by contract No. COO 235–2 of the Atomic Energy Commission.

1. SONNEBORN, T. M., *Proc. Nat. Acad. Sc.,* vol. 46:149 (1960).
2. JACOB, F., and MONOD, J., *in* Cytodifferentiation and Macromolecular Synthesis, (M. Locke, editor), New York, Academic Press, Inc., 30 (1963).
3. NEMER, M., and BARD, S. G., *Science,* vol. 140:664 (1963). NEMER, M., *Biochem. Biophysic. Research Commun.,* vol. 8:511 (1962). See also WILT, F. H., and HULTIN, T., *Biochem. Biophysic. Research Commun.,* vol. 9:313 (1962).
4. BARR, M. L., *Science,* vol. 130:679, 1302 (1959). LYON, M. F., *Nature,* vol. 190:372 (1961). LYON, M. F., *Am. J. Human Genetics,* vol. 14:135 (1962).
5. GALL, J. G., *in* Cytodifferentiation and Macromolecular Synthesis, (M. Locke, editor), New York, Academic Press, Inc., 119 (1963). CALLAN, H. G., *Internat. Rev. Cytol.,* vol. 15:1 (1962). GALL, J. G., and CALLAN, H. G., *Proc. Nat. Acad. Sc.,* vol. 48:562 (1962).
6. IZAWA, M., ALLFREY, V. G., and MIRSKY, A. E., *Proc. Nat. Acad. Sc.,* vol. 49:544 (1963).
7. BEERMANN, W., *Am. Zoologist,* vol. 3:23 (1963).
8. The failure to puff indicates that this gene makes, not a "wrong" product, but no product. In the hybrid the gene puffs in one chromosome of the pair, but remains a condensed band in the other. Hence, failure to puff in the hybrid can hardly be due to absence of the proper stimulus; it is seemingly due to some defect in the response mechanism, presumably in the "operator" on the "operon" hypothesis.
9. CLEVER, U., *Chromosoma,* vol. 12:607 (1961). CLEVER, U., *Developmental Biol.,* vol. 6:73 (1963).
10. MARKERT, C. L., *Science,* vol. 140:1329 (1963). MARKERT, C. L., *in* Cytodifferentiation and Macromolecular Synthesis, (M. Locke, editor), New York, Academic Press, Inc., 65 (1963). MARKERT, C. L., *in* The Nature of Biological Diversity, (J. M. Allen, editor), New York, McGraw-Hill Book Co., Inc., 95 (1963).
11. PUCK, T. T., *in* Rhythmic and Synthetic Processes in Growth, (D. Rudnick, editor), Princeton University Press, 3 (1957). EAGLE, H., *Bact. Rev.,* vol. 22:217 (1958).
12. KELLENBERGER, E., SÉCHAUD, J., and RYTER, A., *Virology,* vol. 8:478 (1959). EPSTEIN, R. H., BOLLÉ, A., STEINBERG, C. M., KELLENBERGER, E., BOY DE LA TOUR, E., CHEVALLEY, R., EDGAR, R. S., SUSMAN, M., DENHARDT, G., and LEILAUSIS, A., *Cold Spring Harbor Symp. Quant. Biol.,* vol. 28:375 (1963).
13. EDGAR, R. S., and LEILAUSIS, A., *Genetics,* vol. 49:649 (1964). EPSTEIN, R. H., STEINBERG, C. M., and BOLLÉ, A., in press.
14. TARTAR, V., The Biology of Stentor, New York, Pergamon Press, 1961.
15. SUZUKI, S., *Bull. Yamagata Univ., Nat. Sc.,* vol. 4:85 (1957).
16. CURTIS, A. S. G., *J. Embryol. Exp. Morph.,* vol. 8:163 (1960), vol. 10:410 (1962).
17. SONNEBORN, T. M., *in* The Nature of Biological Diversity, (J. M. Allen, editor), New York, McGraw-Hill Book Co., Inc., 165 (1963).
18. BEISSON-SCHECROUN, J., *Excerpta Med., Internat. Congr. Series No.* 77, 11 (1964). BEISSON, J., and SONNEBORN, T. M., *Proc. Nat. Acad. Sc.,* vol. 53, Feb. 1965.
19. DIPPELL, R. V., *Excerpta Med., Internat. Congr. Series No.* 77, 16 (1964).
20. WEISS, P., Proc. 5th Conf. on Chemical Research of the Robert A. Welch Foundation, Dec. 4–6, 1961, in press; Weiss, P., *in* The Molecular Control of Cellular Activity, (J. M. Allen, editor), New York, McGraw-Hill Book Co., Inc., 1 (1962).
21. GROBSTEIN, C., *in* Cytodifferentiation and Macromolecular Synthesis, (M. Locke, editor), New York, Academic Press, Inc., 1 (1963).
22. HÄMMERLING, J., *Naturwissenschaften,* vol. 22:829 (1934). BRACHET, J., CHANTRENNE, H., and VANDERHAEGHE, F., *Biochim. et Biophysica Acta,* vol. 18:544 (1955). RICHTER, G., *Biochim. et Biophysica Acta,* vol. 34:407 (1959). OLSZEWSKA, M. J., and BRACHET, J., *Exp. Cell Research,* vol. 22:370 (1961). These papers make no claim for stable m-RNA; all of the later papers report protein synthesis in enucleated cells long after removal

of the nucleus. This can hardly be limited to chloroplasts in view of the regeneration of other cell structures.

23. GIBSON, I., and BEALE, G. H., *Genet. Res. Camb.*, vol. 3:24 (1962); vol. 4:42 (1963); vol. 5:85 (1964).
24. REEVE, E. C. R., and ROSS, G. J. S., *Genet. Res. Camb.*, vol. 3:328 (1962).
25. GIBSON, I., and SONNEBORN, T. M., *Proc. Nat. Acad. Sc.*, vol. 52:869 (1964).
26. STEWARD, F. C., *Science*, vol. 143:20 (1964).
27. BRAUN, A., *Proc. Nat. Acad. Sc.*, vol. 45:932 (1959).
28. SUSSMAN, M., *in* The Chemical Basis of Development, (W. D. McElroy and B. Glass, editors), Baltimore, The Johns Hopkins Press, 264 (1958). BONNER, J. T., The Cellular Slime Molds, Princeton University Press (1959). SUSSMAN, R. R., and SUSSMAN, M., *J. Gen. Microbiol.*, vol. 30:349 (1963).
29. SCHUSTER, F., *J. Protozool.*, vol. 10:297, 313 (1963).
30. SONNEBORN, T. M., *J. Protozool.*, vol. 10:suppl.: 25 (1963).
31. TATUM, E. L., *Science*, vol. 129:1711 (1959).

The Influence of the Environment

1. HUTCHINSON, G. E., *Quart. Rev. Biol.*, vol. 18:331 (1943); with additional data for Sc: BECK, G., *Mikrochemie ver. Mikrochim. Acta*, vol. 34:62 (1948); for Be, B, Co, Hg: FORBES, R. M., COOPER, A. R., and MITCHELL, H. H., *J. Biol. Chem.*, vol. 209:857 (1954); for Si: GETTLER, A. O., and UMBERGER, C. J., *Am. J. Clin. Path.*, Tech. Sect., vol. 9:1 (1945); for F: GUSCHIN, S. K., *Vopr. Pitaniya*, vol. 19:71 (1960); for I: GUSTUN, M. J., *Vopr. Pitaniya*, vol. 18:80 (1959); for As: HERMAN, M. A., WIKTOR, T. J., and VAN HEE, A. A., *Bull. Agr. Congo Belge*, vol. 51:403 (1960); for Br: MORUZZI, G., *Boll. Soc. Ital. Biol. Sper.*, vol. 11:725 (1936); and for Cd: VOĬNAR, A. O., *Akad. Nauk SSSR, Tr. Konf. Mikroelement*, vol. 1950:580 (1952).
2. RYGH, O., *Bull. Soc. Chim. Biol.* (Paris), vol. 31:1052 (1949).
3. REVELLE, R., and SUESS, H. E., *Tellus*, vol. 9:18 (1957). BOLIN, B., and ERIKSSON, E., *in* The Atmosphere and the Sea in Motion, New York, The Rockefeller Institute Press, 130 (1959). ERIKSSON, E., and WELANDER, P., *Tellus*, vol. 8:155 (1956).
4. LIETH, H., *J. Geophys. Research*, vol. 68:3887 (1963). BOLIN, B., and KEELING, C. D., *J. Geophys. Research*, vol. 68:3899 (1963).
5. CALLENDAR, G. S., *Tellus*, vol. 10:243 (1958). BOLIN, B., and KEELING, C. D., *J. Geophys. Research*, vol. 68:3899 (1963).
6. HARDIN, G., *Science*, vol. 131:1292 (1960).
7. RILEY, G. A., *in Marine Biol. 1*, Am. Inst. Biol. Sc., 70 (1963). HUTCHINSON, G. E., *Am. Naturalist*, vol. 95:137 (1961).
8. RICHTER, C. P., *Proc. Nat. Acad. Sc.*, vol. 46:1506 (1960).
9. COWGILL, U. M., BISHOP, A., ANDREW, R. J., and HUTCHINSON, G. E., *Proc. Nat. Acad. Sc.*, vol. 48:238 (1962). HARRISON, J. L., *Bull. Raffles Mus.*, vol. 24:109 (1952).
10. ALEXANDER, R. D., and MOORE, T. E., *Misc. Publ. Mus. Zool., Univ. Mich.*, vol. 121:1 (1962). DYBAS, H. S., and LLOYD, M., *Ecology*, vol. 43:444 (1962).
11. ASHMULE, N. P., *Ibis*, vol. 103b:458 (1963).

Physiological and Cultural Determinants of Behavior

The author wishes to thank Drs. W. Goodenough, G. P. Murdock, C. Pfaffman, and H. Riecken for making helpful comments on a rough draft of this paper, and Mrs. Emily Derow for overtime work in preparing the final copy in time for the meeting.

Some of the work on this paper was supported by grant MY 647 from NIMH of the United States Public Health Service.

1. ALMOND, G. A., and VERBA, S., The Civic Culture, Princeton University Press (1963).
2. ANDERSSON, B., EKMAN, L., GALE, C. C., and SUNDSTEN, J. W., Activation of the thyroid gland by cooling of the preoptic region in the goat, *Acta Physiol. Scand.*, vol. 54:191–192 (1962).
3. ANDERSSON, B., and LARSSON, B., Influence of local temperature changes in the preoptic area and rostral hypothalamus on the regulation of food and water intake, *Acta Physiol. Scand.*, vol. 52:75–89 (1961).
4. ARORA, H. L., and SPERRY, R. W., Optic nerve regeneration after cross union of medial and lateral optic tracts, *Am. Zool.*, vol. 2:389 (1962).
5. ATTARDI, D. G., and SPERRY, R. W., Preferential

selection of central pathways by regenerating optic fibers, *Exp. Neurol.,* vol. 7:46–64 (1963).
6. Berlyne, D. E., Conflict Arousal and Curiosity, New York, McGraw-Hill Publishing Co., Inc. (1960).
7. Boas, F., The Central Eskimo, Bureau of American Ethnology, Annual Report, vol. 6 (1888).
8. Bowlby, J., Maternal Care and Mental Health, Geneva, World Health Organization (1952).
9. Broadbent, D. E., Perception and Communication, New York, London, Pergamon Press (1958).
10. Bronfenbrenner, U., Soviet methods of character education; some implications for research, *Am. Psychol.,* vol. 17:550–564 (1962).
11. Bush, R. R., Galanter, E., and Luce, R. D., Handbook of Mathematical Psychology, New York, John Wiley and Sons, Inc. (1963).
12. Campbell, A., Miller, W., and Stokes, E., The American Voter, New York, John Wiley and Sons, Inc. (1960).
13. Child, V. G., What Happened in History, Harmondworth, England, Penguin Books (1948).
14. Cofer, C. N., Verbal Learning and Verbal Behavior, New York, McGraw-Hill Publishing Co., Inc. (1961).
15. Coleman, J. S., The Adolescent Society, Glencoe, The Free Press (1961).
16. Delafresnaye, J. F., Brain Mechanisms and Learning, Oxford, Blackwell Scientific Publications (1961).
17. Dollard, J., and Miller, N. E., Personality and Psychotherapy, New York, McGraw-Hill Publishing Co., Inc. (1950).
18. Eccles, J. C., The Physiology of Synapses, Heidelberg, Springer Verlag and New York, Academic Press, Inc. (1964).
19. Fields, W. S., and Abbott, W., Information Storage and Neural Control, Springfield, Illinois, Charles C Thomas (1963).
20. Fisher, A., Maternal and sexual behavior induced by intracranial chemical stimulation, *Science,* vol. 124:228–229 (1956).
21. French, J. D., Porter, R. W., Cavanaugh, E. B., and Longmire, R. L., Experimental observations on "psychosomatic" mechanisms. I. Gastrointestinal disturbances, *Arch. Neurol. and Psychiat.,* vol. 72:267–281 (1954).
22. Gaddum, J. H., Chemical transmission in the central nervous system, *Nature,* vol. 197:741–743 (1963).
23. Gewirtz, J. L., and Baer, D. M., The effect of brief social deprivation on behavior for a social reinforcer, *J. Abnorm. Soc. Psychol.,* vol. 56:49–56 (1958).
24. Gibson, E. J., Perceptual learning, *in* Annual Review of Psychology, (A. R. Farnsworth, editor), Palo Alto, California, Annual Reviews, Inc., vol. 14 (1963).
25. Grossman, S. P., Direct adrenergic and cholinergic stimulation of hypothalamic mechanism, *Am. J. Physiol.,* vol. 202:872–882 (1962).
26. Grossman, S. P., Effects of adrenergic and cholinergic blocking agents on hypothalamic mechanisms, *Am. J. Physiol.,* vol. 202:1230–1236 (1962).
27. Gunn, C. G., Friedman, M., and Byers, S. O., Effects of chronic hypothalamic stimulation upon cholesterol-induced atherosclerosis in the rabbit, *J. Clin. Inv.,* vol. 39:1963–1972 (1960).
28. Hagen, E. E., On the Theory of Social Change, Dorsey Press (1962).
29. Harlow, H. F., and Zimmerman, R. R., Affectional responses in the infant monkey, *Science,* vol. 130:421–432 (1959).
30. Harris, G. W., and Levine, S., Sexual differentiation of the brain and its experimental control, *J. Physiol.,* vol. 163:42–43 (1962).
31. Henry, J., Jungle People, New York, J. J. Augustin Press (1941).
32. Holland, J. G., and Skinner, B. F., The Analysis of Behavior, New York, McGraw-Hill Publishing Co., Inc. (1961).
33. Hollingshead, A. B., and Redlich, F., Social Class and Mental Illness, New York, John Wiley and Sons, Inc. (1958).
34. Hubel, D. H., and Wiesel, T. N., Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol.,* vol. 160:106–154 (1963).
35. Hummel, H. T., Hardy, J. D., and Fusco, M. M., The thermoregulatory response to hypothalamic cooling in unanesthetized dogs, *Am. J. Physiol.,* vol. 198:481–486 (1960).
36. Hutchinson, G. E., "The Influence of the Environment," this volume.
37. Kimble, G. A., Hilgard and Marquis' Conditioning and Learning, New York, Appleton-Century-Crofts, Inc., 2nd edition (1961).
38. Landauer, T. K., and Whiting, J. W. M., Infantile Stimulation and Adult Stature of

Human Males, *Am. Anthropol.,* vol. 66:1007 (1964).

39. LEIGHTON, A. H., My Name is Legion, New York, Basic Books, Inc. (1959).
40. LEVINE, S., Psychophysiological effects of infant stimulation, *in* Roots of Behavior, (E. L. Bliss, editor), New York, Paul B. Hoeber, Inc., 246–253 (1962).
41. LIVINGSTON, R. B., Central control of receptors and sensory transmission systems, *in* Handbook of Physiology, (J. Field, H. W. Magoun, and V. E. Hull, editors), Baltimore, The Williams and Wilkins Co., Vol. 1, section 1, Neurophysiology (1959).
42. LOGAN, F. A., Incentive: How the Conditions of Reinforcement Affect the Performance of Rats, New Haven, Yale University Press (1960).
43. MAGOUN, H. W., The Waking Brain, Springfield, Illinois, Charles C Thomas, 2nd edition (1963).
44. MALINOWSKI, B., Argonauts of the Western Pacific, New York, E. P. Dutton & Co., 63–64 (1922).
45. MAYR, E., "The Evolution of Living Systems," this volume.
46. MILLER, N. E., Learning and performance motivated by direct stimulation of the brain, *in* Electrical Stimulation of the Brain, University of Texas Press, 387–396 (1961).
47. MILLER, N. E., Some reflections on the law of effect produce a new alternative to drive reduction, *in* Nebraska Symposium on Motivation for 1963, University of Nebraska Press, 65–112 (1963).
48. MILLER, N. E., Some motivational effects of electrical and chemical stimulation of the brain, *EEG Journal*, suppl. 24, Elsevier, 247–259 (1963).
49. MILLER, N. E., Animal experiments on emotionally-induced ulcers, in *Proc. World Congress Psychiatry,* Montreal, Canada, June 4–10, 1961, Toronto, University of Toronto Press, vol. 3:213–219 (1963).
50. MILLER, N. E., Some implications of modern behavior theory for personality and psychotherapy, *in* Personality Change, (D. Byrne and P. Worchel, editors), New York, John Wiley and Sons, Inc. (1964).
51. MILLER, N. E., Certain recent developments in experimental psychology, *Proc. Roy. Soc. London, Series B,* vol. 158:481–497 (1963).
52. MILLER, N. E., GOTTESMAN, K. S., and EMERY, N., Dose-response to carbachol and norepinephrine in rat's hypothalamus, *Am. J. Physiol.,* vol. 206:1384–1388 (1964).
53. MILLER, N. E., PFAFFMAN, C., and SCHLOSBERG, H., Aspects of psychology and psychophysiology in the U.S.S.R., *in* Some Views on Soviet Psychology, (R. A. Bauer, editor), New York, Timely Press, Inc., 226–229 (1962).
54. MURDOCK, G. P., Our Primitive Contemporaries, New York, The Macmillan Co., 110 (1934).
55. MURDOCK, G. P., Social Structure, New York, The Macmillan Co. (1949).
56. MYERS, J. K., and ROBERTS, B. H., Family and Class Dynamics in Mental Illness, New York, John Wiley and Sons, Inc. (1959).
57. NAKAYAMA, T., HAMMEL, H. T., HARDY, J. D., and EISENMAN, J. S., Thermal stimulation of electrical activity of single units of the preoptic region, *Am. J. Physiol.,* vol. 204:1122–1126 (1963).
58. NOVIN, D., The relation between electrical conductivity of brain tissue and thirst in the rat, *J. Comp. Physiol. Psychol.,* vol. 55:145–154 (1962).
59. POLISH, E., BRADY, J. V., MASON, J. W., THACH, J. S., and NIEMACK, W., Gastric contents and the occurrence of duodenal lesions in the rhesus monkey during avoidance behavior, *Gastroenterology,* vol. 43:193–201 (1962).
60. PYE, L., Politics, Personality and Nation Building: Burma's Search for Identity, New Haven, Yale University Press (1962).
61. REYNOLDS, L., Economics, Homewood, Illinois, Richard D. Irwin, Inc. (1963).
62. SARASON, S. B., and GLADWIN, T., Psychological Problems in Mental Deficiency: Part II, New York, Harper & Brothers (1959).
63. SAWYER, C. H., and KAWAKAMI, M., Interaction between the central nervous system and hormones influencing ovulation, *in* The Control of Ovulation, (C. A. Villee, editor), New York, Pergamon Press, 79–97 (1961).
64. SCOTCH, N. A., Sociocultural factors in the epidemiology of Zulu hypertension, *Am. J. Pub. Health,* vol. 53:1205–1213 (1963).
65. SCOTT, J. P., Critical periods in behavioral development, *Science,* vol. 138:949–957 (1962).
66. SONNEBORN, T. M., "The Differentiation of Cells," this volume.
67. SPENCER, W., and GILLEN, F., The Arunta,

London, The Macmillan Co., vol. 1:27–28 (1927).
68. SROLE, L., Mental Health in the Metropolis: Midtown Manhattan Study, New York, McGraw-Hill Publishing Co., Inc. (1962).
69. STAMLER, R., Acculturation and Negro blue-collar workers, *in* The Blue Collar World, (A. B. Shostak and W. Gomberg, editors), New Jersey, Prentice-Hall, Inc., 282–298 (1964).
70. WARNER, W. L., MEEKER, M., and ELLIS, K., Social Class in America: A Manual for the Measurement of Social Status, Chicago, Science Research Associates (1949).
71. WILSON, P., The Tsimihety and the Sihanaka, Human Relations Area Files, New Haven, personal communication.

THE SCIENTIFIC ENDEAVOR [pages 267–309]

Synthesis and Applications of Scientific Knowledge for Human Use

1. George Parsons Lathrop, "Talks with Edison," *Harper's New Monthly Magazine*, LXXX [1890], 432–34.
2. *Nature*, XVIII, pp. 159–163, June 6 [1878].

Illustrations

HISTORY OF THE UNIVERSE

PART TITLE Courtesy of Harvard College Observatory.

The Origin of the Elements

FIGURES 3, 4 Courtesy of Mount Wilson-Palomar Observatories.
FIGURE 12 The spectrum of R. Andromedae was obtained by P. W. Merrill, and the upper two spectra by E. M. and G. R. Burbidge.

The History of Stars and Galaxies

FIGURE 1 After Chushiro Hayashi, Department of Nuclear Science, Kyoto University, Kyoto, Japan.
FIGURE 2 Courtesy of Mount Wilson-Palomar Observatories.
FIGURE 3 After Chushiro Hayashi, Department of Nuclear Science, Kyoto University, Kyoto, Japan.
FIGURE 4, 5 Courtesy of Allan Sandage.
FIGURE 6 Courtesy of O. C. Wilson.
FIGURE 7, 8 Courtesy of the Mount Wilson-Palomar Observatories.
FIGURE 9 Courtesy of Mount Wilson-Palomar Observatories. Charles R. Lynds and Allan Sandage.
FIGURE 10 Courtesy of Mount Wilson-Palomar Observatories.

The History of the Solar System

FIGURE 6 According to F. Hoyle.
FIGURE 8 After W. C. DeMarcus.
FIGURE 12 Courtesy of Harvard College Observatory.

The Origins of Life

"Ammonia! Ammonia!" Drawing by R. Grossman; © 1962 The New Yorker Magazine, Inc.

NATURE OF MATTER

PART TITLE The first published report on this work appeared in the *Physical Review Letters,* volume 12, page 204, February 1964. Photograph courtesy of Brookhaven National Laboratory.

The Organization of Living Matter

FIGURE 1 From Michaelis, L., *Arch. mikroskop. Anat. Entwicklungsmechn.,* volume 55, page 558 (1900), Figure 6.
FIGURE 3 Courtesy of Dr. Walther Stoeckenius, The Rockefeller Institute.
FIGURE 4 Mathews, A., The changes in structure of the pancreas cell, *J. Morphol.,* volume 15, suppl., page 171 (1899), Figure 16.
FIGURE 14 Courtesy of Dr. D. J. L. Luck, The Rockefeller Institute.
FIGURE 15 Courtesy of Dr. A. Tomasz and Dr. J. Jamieson, The Rockefeller Institute.

THE DETERMINANTS AND EVOLUTION OF LIFE

The Differentiation of Cells

FIGURE 1*a* Redrawn after Gall, J. G., Chromosomes and cytodifferentiation, *in* Cytodifferentiation and Macromolecular Synthesis, (M. Locke, editor), New York, Academic Press, Inc., 1963, page 119, Figure 3.
FIGURE 1*b* Redrawn after Beermann, W., Cytological aspects of information transfer in cellular differentiation, *Am. Zoologist,* volume 3, page 23 (1963), Figure 1.

THE SCIENTIFIC ENDEAVOR

PART TITLE Courtesy Life Magazine © copyright Time Inc.

PART TITLE *A Century of Scientific Conquest* courtesy of Frank R. Fisher.