



# Introduction to Concepts and Theories in Physical Science

## Citation

Holton, Gerald. Introduction to Concepts and Theories in Physical Science. Reading, MA: Addison-Wesley, 1952.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:41650112>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

INTRODUCTION TO  
CONCEPTS AND  
THEORIES IN  
PHYSICAL SCIENCE

This book is in the  
ADDISON-WESLEY SERIES IN PHYSICS

INTRODUCTION TO  
CONCEPTS AND  
THEORIES IN  
PHYSICAL SCIENCE

GERALD HOLTON, Ph.D.

*Professor of Physics*  
*Harvard University*



ADDISON-WESLEY PUBLISHING COMPANY, INC.  
READING, MASSACHUSETTS, U. S. A.  
LONDON, ENGLAND



*Copyright © 1952*

**ADDISON-WESLEY PUBLISHING COMPANY, INC.**

---

*Printed in the United States of America*

ALL RIGHTS RESERVED. THIS BOOK, OR PARTS THERE-  
OF, MAY NOT BE REPRODUCED IN ANY FORM WITHOUT  
WRITTEN PERMISSION OF THE PUBLISHERS.

*Library of Congress Catalog No. 52-7666*

*Sixth printing—January, 1962*

TO E.C.K.  
COLLEAGUE AND TEACHER



# CONTENTS

Introduction . . . . .	xiii
------------------------	------

## PART A: THE STUDY OF MOTION

CHAPTER 1. VELOCITY AND ACCELERATION . . . . .	1
1-1 Constant velocity . . . . .	1
1-2 The concept of average velocity . . . . .	6
1-3 Instantaneous velocity . . . . .	7
1-4 Acceleration . . . . .	9
1-5 Equations of motion for constant acceleration . . . . .	11
CHAPTER 2. GALILEO AND THE KINEMATICS OF FREE FALL . . . . .	17
2-1 Introduction . . . . .	17
2-2 Aristotelian physics . . . . .	18
2-3 Galileo. Biographical sketch . . . . .	23
2-4 Galileo's <i>Two New Sciences</i> . . . . .	25
CHAPTER 3. PROJECTILE MOTION . . . . .	36
3-1 Projectile with initial horizontal motion . . . . .	36
3-2 Introduction to vectors . . . . .	42
3-3 The general case of projectile motion . . . . .	44
3-4 Applications of the law of projectile motion . . . . .	49
3-5 Galileo's conclusions . . . . .	50
3-6 Summary . . . . .	52

## PART B: THE STUDY OF FORCES

CHAPTER 4. NEWTON'S LAWS OF MOTION . . . . .	57
4-1 Newton's first law of motion . . . . .	57
4-2 Newton's second law of motion . . . . .	59
4-3 Standard of mass . . . . .	62
4-4 Weight . . . . .	63
4-5 The equal-arm balance . . . . .	67
4-6 Systems of units . . . . .	69
4-7 Examples and applications of Newton's second law of motion . . . . .	72
4-8 Newton's third law of motion . . . . .	78
4-9 Examples and applications of Newton's third law . . . . .	79
CHAPTER 5. ROTATIONAL MOTION . . . . .	85
5-1 Kinematics of constant circular motion . . . . .	85
5-2 Vector algebra . . . . .	88
5-3 Centripetal acceleration . . . . .	93
5-4 Centripetal force . . . . .	95

5-5	Uniformly accelerated circular motion . . . . .	97
5-6	Simple harmonic motion . . . . .	99
5-7	The case of the simple pendulum . . . . .	102
5-8	The vibrating string . . . . .	104
5-9	Wave propagation . . . . .	106
5-10	Conclusion . . . . .	107
<b>PART C: THE LAWS OF PLANETARY MOTION AND UNIVERSAL GRAVITATION</b>		
CHAPTER 6.	THE ASTRONOMY OF ANCIENT GREECE . . . . .	113
6-1	Plato's problem . . . . .	114
6-2	The Aristotelian system . . . . .	116
6-3	The heliocentric theory . . . . .	118
6-4	Modified geocentric theories . . . . .	121
6-5	The success of the Ptolemaic system . . . . .	124
CHAPTER 7.	COPERNICUS' HELIOCENTRIC THEORY . . . . .	127
7-1	Europe reborn . . . . .	127
7-2	The Copernican system . . . . .	127
7-3	Bracing the system . . . . .	132
7-4	The opposition to Copernicus' theory . . . . .	133
7-5	Historic consequences . . . . .	136
CHAPTER 8.	THE NATURE OF SCIENTIFIC THEORY . . . . .	138
8-1	The purpose of theories . . . . .	138
8-2	Criteria for a good theory in physical science . . . . .	142
CHAPTER 9.	KEPLER'S LAWS . . . . .	148
9-1	Kepler's first law . . . . .	150
9-2	Kepler's second law . . . . .	154
9-3	The meaning of the second law . . . . .	155
9-4	Kepler's third law . . . . .	156
9-5	The new concept of physical law . . . . .	157
CHAPTER 10.	GALILEO'S CONTRIBUTION TO ASTRONOMY . . . . .	160
10-1	The telescopic evidences . . . . .	161
10-2	Science and freedom . . . . .	165
CHAPTER 11.	NEWTON'S LAW OF UNIVERSAL GRAVITATION . . . . .	169
11-1	Science in the seventeenth century . . . . .	169
11-2	A short sketch of Newton's life . . . . .	171
11-3	Newton's <i>Principia</i> . . . . .	173
11-4	Derivation of the law of universal gravitation . . . . .	175
11-5	Some influences on Newton's work . . . . .	188
11-6	Some consequences of the law of universal gravitation . . . . .	190
11-7	"I frame no hypotheses" . . . . .	205
11-8	Newton's place in modern science . . . . .	208

PART D: ON STRUCTURE AND METHOD  
IN PHYSICAL SCIENCE

CHAPTER 12. CONCEPTS . . . . .	215
12-1 Introduction: the search for constancies in change . . . . .	215
12-2 Science and nonscience . . . . .	216
12-3 The lack of a single method . . . . .	218
12-4 Physical concepts: measurement and definition . . . . .	221
12-5 Primary and secondary quantities . . . . .	226
12-6 Mathematical law and abstraction . . . . .	227
12-7 Explanation . . . . .	231
CHAPTER 13. THE GROWTH OF SCIENCE . . . . .	234
13-1 The free license of creativity . . . . .	234
13-2 The natural selection of physical concepts . . . . .	237
13-3 Motivation . . . . .	240
13-4 Objectivity . . . . .	244
13-5 Fact and interpretation . . . . .	246
13-6 How science grows . . . . .	248
13-7 Consequences of the model . . . . .	250
CHAPTER 14. LAWS . . . . .	257
14-1 Opinions on scientific procedure . . . . .	257
14-2 A sequence of elements in formulations of laws . . . . .	263
14-3 The limitations of physical law . . . . .	269
14-4 The content of science. Summary . . . . .	271

PART E: THE LAWS OF CONSERVATION

CHAPTER 15. THE LAW OF CONSERVATION OF MATTER . . . . .	279
15-1 Prelude to the conservation law . . . . .	279
15-2 Steps toward a formulation . . . . .	280
15-3 Lavoisier's experimental proof . . . . .	282
CHAPTER 16. THE LAWS OF CONSERVATION OF MOMENTUM . . . . .	286
16-1 Introduction . . . . .	286
16-2 Definition of momentum . . . . .	288
16-3 The law of conservation of (linear) momentum . . . . .	290
16-4 Examples involving collisions . . . . .	291
16-5 Examples involving explosions . . . . .	294
16-6 Further examples . . . . .	296
16-7 Rotational inertia . . . . .	303
16-8 Angular momentum . . . . .	307
16-9 The law of conservation of angular momentum . . . . .	308
16-10 Conclusion . . . . .	311
CHAPTER 17. THE LAW OF CONSERVATION OF ENERGY . . . . .	313
17-1 Vis viva . . . . .	313
17-2 Preliminary questions. The pile driver . . . . .	317

17-3	The concept of work . . . . .	319
17-4	Various forms of energy . . . . .	321
17-5	The conservation law, first form; and applications . . . . .	324
17-6	First extension of the conservation law . . . . .	330
17-7	Further extension and applications of the LCE . . . . .	332
17-8	The mechanical equivalent of heat. Mayer and Joule . . . . .	345
17-9	General illustrations of the LCE . . . . .	357

**PART F: ORIGINS OF THE ATOMIC THEORY IN PHYSICS  
AND CHEMISTRY**

CHAPTER 18.	THE PHYSICS OF GASES AND OF HEAT . . . . .	367
18-1	The nature of gases—early concepts . . . . .	367
18-2	Digression: the general gas law . . . . .	370
18-3	Two gas models . . . . .	373
18-4	The caloric theory of heat . . . . .	376
CHAPTER 19.	THE ATOMIC THEORY OF CHEMISTRY . . . . .	385
19-1	Introduction . . . . .	385
19-2	Dalton's model of gases . . . . .	387
19-3	The properties of Dalton's chemical atom . . . . .	390
19-4	The early achievements of Dalton's theory . . . . .	395
19-5	Gay-Lussac's law of combining volumes of reacting gases . . . . .	398
19-6	Avogadro's hypothesis . . . . .	401
19-7	Chemistry after Avogadro . . . . .	406
19-8	Valence . . . . .	412
19-9	The search for regularity among the list of elements . . . . .	415
19-10	The early periodic table of elements . . . . .	418
19-11	Consequences of the periodic law . . . . .	425
19-12	The modern periodic table . . . . .	427
CHAPTER 20.	THE KINETIC-MOLECULAR THEORY OF GASES . . . . .	432
20-1	Introduction . . . . .	432
20-2	Some qualitative successes of the kinetic-molecular theory . . . . .	434
20-3	Model of a gas, and assumptions in the kinetic theory . . . . .	436
20-4	The derivation of the pressure formula . . . . .	443
20-5	Consequences and verification of the kinetic theory . . . . .	447

**PART G: QUANTUM THEORY AND THE NUCLEAR ATOM**

CHAPTER 21.	ELECTROSTATICS . . . . .	473
21-1	Introduction . . . . .	473
21-2	Electrification by friction . . . . .	473
21-3	Law of conservation of charge . . . . .	474
21-4	Modern model for electrification . . . . .	475
21-5	Insulators and conductors . . . . .	476

21-6	The electroscope . . . . .	480
21-7	The quantity of charge . . . . .	482
21-8	Coulomb's law of electrostatics . . . . .	483
21-9	Applications of Coulomb's law . . . . .	485
21-10	The electrostatic field . . . . .	486
21-11	Lines of force . . . . .	489
21-12	Electric potential difference—qualitative . . . . .	490
21-13	Potential difference—quantitative . . . . .	492
21-14	Uses of the concept of potential . . . . .	494
21-15	Electrochemistry . . . . .	496
21-16	Atomicity of charge . . . . .	498
CHAPTER 22. PROPERTIES OF LIGHT . . . . .		501
22-1	Introduction . . . . .	501
22-2	Electromagnetic waves and ether . . . . .	501
22-3	Hertz's experiments . . . . .	508
22-4	Continuous emission spectra . . . . .	511
22-5	Line emission spectra . . . . .	515
22-6	Absorption line spectra . . . . .	517
22-7	Balmer's formula . . . . .	523
CHAPTER 23. THE QUANTUM THEORY OF LIGHT . . . . .		528
23-1	Planck's empirical emission formula . . . . .	528
23-2	The quantum hypothesis . . . . .	530
23-3	The photoelectric effect . . . . .	537
23-4	Einstein's photon theory . . . . .	540
23-5	The photon-wave dilemma . . . . .	545
23-6	Applications of the photon concept . . . . .	547
23-7	Quantization in science . . . . .	550
CHAPTER 24. RADIOACTIVITY AND THE NUCLEAR ATOM . . . . .		552
24-1	Isotopes . . . . .	552
24-2	Radioactive "half life" . . . . .	556
24-3	Radioactive series . . . . .	559
24-4	Rutherford's nuclear model . . . . .	562
24-5	Moseley's x-ray spectra . . . . .	571
24-6	Further concepts of nuclear structure . . . . .	575
CHAPTER 25. BOHR'S MODEL OF THE ATOM . . . . .		582
25-1	Introduction . . . . .	582
25-2	Energy levels in hydrogen atoms . . . . .	583
25-3	Correspondence principle . . . . .	590
25-4	The explanation of hydrogen line emission spectra . . . . .	593
25-5	Further results in Bohr's original paper . . . . .	597
25-6	Further developments . . . . .	607
25-7	Epilogue . . . . .	622



APPENDIX I. FUNDAMENTAL CONSTANTS . . . . .	625
APPENDIX II. TABLE OF CONVERSION FACTORS . . . . .	627
APPENDIX III. ALPHABETIC LIST OF THE ELEMENTS . . . . .	629
APPENDIX IV. PERIODIC TABLE OF THE ELEMENTS . . . . .	631
APPENDIX V. SUMMARY OF SOME TRIGONOMETRIC RELATIONS . . . . .	633
APPENDIX VI. NATURAL TRIGONOMETRIC FUNCTIONS . . . . .	637
APPENDIX VII. COMMON LOGARITHMS . . . . .	638
ANSWERS TO PROBLEMS . . . . .	641
INDEX . . . . .	643

## INTRODUCTION

In the last decade or two, the function and form of introductory science courses in our colleges have been the subject of general discussion. We have already begun to see some stimulating consequences of this widespread interest in re-analysis and re-evaluation, principally the inauguration of general education courses in their several forms.

In this book I have attempted to incorporate some of the conclusions drawn from an illuminating experience, that of teaching simultaneously for a number of years a conventional type of introductory college physics course and also in general education courses in physical science. Perhaps the novelty of some features of the book will excuse a few introductory remarks concerning the guiding ideas that have shaped it.

The old postulate has remained: the principal task for an introductory course in the physical sciences must surely be a clear or, rather, a dedicated presentation of the fundamental concepts and theories. A firm knowledge of fundamentals is the basis for whatever is to be accomplished in addition. Moreover, by presenting neither an encyclopedic account of all details nor a survey treatment, by concentrating instead on the meaning and power of some basic ideas—even though there can probably be no general agreement on just what topics one should include—one may hope to leave the student with a view of physical science as an interconnected achievement; one can engage his attention in those lengthier, more rigorous, penetrating arguments and derivations without which an honest understanding of the processes of science is hardly attainable.

While therefore the focus in this book is on the subject matter of physical science, there are also other aims—above all the presentation of science as *experience*, as an integrated and exciting intellectual adventure. Alfred North Whitehead, in the preface of his collection of essays, *The Aims of Education*, protested “against dead knowledge, that is to say, against inert ideas.” Indeed, the record would tend to show that in the past it has often been found difficult to avoid the cataloguing of inert ideas, particularly in science textbooks. This does not refer to the text directed at those students who concentrate or major in one of the physical sciences; they bring to their studies a commitment and a motivation that give life to the basic concepts and

theories; furthermore, they can look forward to a fulfillment in later courses of all promises implicit in the introductory treatment. But this commitment, motivation, and long-range interest is much attenuated in the student who does not plan to devote his life to physical science, for whom the first and perhaps the last formal contact with the subject is a single course. Yet he too is eager to understand the achievements and methods of science. He responds warmly to sympathetic guidance.

His dilemma was well analyzed in the searching Report on *General Education in a Free Society* (Harvard University Press, 1945, pp. 220-222):

“From the viewpoint of general education the principal criticism to be leveled at much of present college instruction in science is that it consists of courses in special fields, directed toward training the future specialist and making few concessions to the general student. Most of the time in such courses is devoted to developing a technical vocabulary and technical skills, and to a systematic presentation of the accumulated fact and theory which the science has inherited from the past. Comparatively little serious attention is given to the examination of basic concepts, the nature of the scientific enterprise, the historical development of the subject, its great literature, or its interrelationships with other areas of interest and activity. What such courses frequently supply are only the bricks of the scientific structure. The student who goes on into more advanced work can build something from them. The general student is more likely to be left simply with bricks. Eventually he constructs his educational edifice elsewhere with other materials. . . .

“The body of science includes not only special knowledge and skills but conceptual interrelations, a world-view, and a view of the nature of man and knowledge, which together constitute the philosophy of science; a history which forms a continuous and important segment of all human history; and writings which include some of the most significant and impressive contributions to all literature. These aspects of the sciences are frequently almost entirely neglected in the college teaching of science . . .”

Certainly, on their own merits, the inclusion of history and of philosophy of science is justified in an introductory course for the nonspecialist. But beyond that, their use in this book has a further threefold purpose. First, they prepare the appropriate setting in

which a particular idea came to have meaning and importance. Second, they provide insight into the sources, motivations, and methods of approach of the founders of science, illustrating the human triumph behind the bare abstraction. And third, they present science as one facet of the great quest for knowledge which our free society holds to be a most cherished right and a most respected driving force. At the end of such a course, the student will be in a position to know the principal laws and the evolution of key conceptual schemes, but it is then also hoped that he, as a responsible citizen, will understand the criteria of validity in scientific thought, the conditions that aid the fruitful growth of science—and perhaps even the exhilaration that binds his instructor to the scientific profession.

A few specific remarks may now be made about the form of presentation. It was not dictated by any single pedagogic method. The sequence follows an eclectic pattern, most frequently a series of modified case accounts involving the great conceptual schemes. But while the latter are analyzed, as far as practical, in terms of original source materials, a contemporary mode of exposition has been adopted instead of strictly historical and methodological ones wherever these would not aid materially in clarifying the fundamental ideas.

Being directed primarily to students not planning to become professional physicists or chemists, the book will find use in two rather different contexts: the course explicitly of the general education type and the liberal arts physics course for nonscience majors. As to the first of these, I have in mind particularly a course analogous in many ways to the one described in the Report on General Education in the following words (*op. cit.*, pp. 226–227):

“. . . Rather than provide the student with a systematic presentation of the materials of one science, this course should develop particular aspects of the scientific enterprise within the whole range of the physical sciences. To give the course greater unity, it should be built about a core of physics. Materials from other sciences—chemistry, astronomy, and geology—should be introduced only to the degree that they are pertinent to the problems under discussion. The course would probably omit, for example, descriptive chemistry and descriptive astronomy. It should, however, explore basic chemical concepts: atomic theory, the periodic system, laws of chemical combination, valence, and so on. Similarly, celestial mechanics might provide the material for much of its discussion of dynamical principles.

“Such a course must discard at the outset any attempt to survey the material of the sciences which compose it. Rather, it must look to some dominant intellectual pattern to guide its selection of material. In the present case the pattern is to be found in the development of basic physical principles and concepts, and the methods and approaches by which they have been developed.

“This is not intended to be merely a course about science. It will contain much solid scientific content. The student will learn fundamental facts and laws and will solve problems theoretically and in the laboratory. He will do so, however, with a highly selected subject matter, which in every case is chosen to subserve the major aims of the course.

“The emphasis on historical development in this course is in no sense to constitute merely a humanistic garnishing of its factual material. On the contrary, it is introduced to illuminate and vitalize the content with which it is integrated. The attempt should be made in this course to teach science as part of the total intellectual and historical process, of which, in fact, it has always been an important part. The student should gain thereby an insight into the principles of science, and appreciation of the values of the scientific enterprise; and he should also learn much of the subject matter of the physical sciences.”

The second group to which this book is directed is that of students in the liberal arts course in physics for whom the further curriculum does not offer another opportunity to obtain the general education background. Here the instructor may possibly wish to omit such chapters as do not primarily involve physics. Or he may supplement this selected account with whatever topics he chooses from an orthodox type of physics text, in accord with the particular needs of his class; I myself have used preliminary editions of this material in conjunction with such a text in an introductory physics course attended largely by premedical students. I am inclined to agree with the growing opinion that the influence of general education programs upon the teaching of just such groups will eventually be counted as one of its most significant contributions to the educational pattern.

As reference material, these studies should also aid in teaching the strictly technical course for physics majors, for whom the Report on General Education pleads (*op. cit.*, p. 221):

“. . . It frequently happens that even the student who concentrates in a science is preoccupied with his specialty to such a

degree that he fails to achieve a view of science as a whole and of the interrelationships of the special fields within it. A general education in science needs to be provided for the future scientist or technologist as well as for the general student . . .”

The general plan of this book may be outlined as follows: The twenty-five chapters are divided into seven Parts. The first of these, Part A, introduces physical science in its historic context by the study of *motion*. After preliminary consideration of simple kinematics, definitions of terms, and a review of elementary mathematical methods, the discussion turns to an examination of Galileo's work on free fall and on general projectile motion, although both cases are carried through beyond the original scope of investigation. It is intended that the detailed attention to projectile motion will create early proficiency in the use of mathematical law, and confidence in its power.

Part B continues to prepare the foundation of physical science by a study of *forces*. Newton's laws are analyzed, and the methods are developed for dealing with problems in the dynamics of rectilinear, rotational, and vibrational motion.

In Part C, the tools, concepts, and theorems discussed in previous chapters are applied to the ancient problem of *planetary motion* and to an analysis of *Newton's law of universal gravitation*. The Ptolemaic and Copernican theories of the solar system are examined in some detail. After an introductory consideration of the nature of scientific theory based on the foregoing cases, Kepler's laws and Galileo's contributions to astronomical theory are assessed, leading to a detailed study of the genesis and power of Newton's crowning achievement.

An examination involving material from the preceding parts is then made in Part D of *the structure and methods of physical science*, i.e., the role in the growth of science of concepts and hypothesis, mathematics and imaginative insight, theory and experiment.

With Part E, the text turns to the final problem of mechanics, namely, *the laws of conservation*—of mass, of momentum, and of energy. The account of the work of Lavoisier, Huygens, Mayer, and Joule is then put on a contemporary basis by illustrations of the wide applicability of these laws to the general problems of physical science.

Presenting the *origins of the atomic theory in physics and chemistry*, Part F includes the early physics of gases and heat, and leads naturally by way of the work of Dalton and Avogadro, to a consideration of Mendeléeff's periodic table. After applications of the atomic the-

ory of chemistry have been made, the atomic theory of physics is discussed, with main emphasis upon the development of the kinetic-molecular theory of gases at the hands of Joule, and leading, beyond that, to the study of specific heats.

The classical model of the atom is finally replaced in Part G, which is devoted to a study of *quantum theory* and *the nuclear atom*. Here the fundamentals of electricity are investigated first, followed by a consideration of the properties of light and, with these, of the work of Maxwell, Hertz, and the 19th century spectroscopists. The physics of Planck's quantum theory and Einstein's photon theory is presented, and from there the historical development leads to radioactivity and Rutherford's work on the nuclear atom. Finally, Bohr's model of the atom is studied in detail, including its successes in gathering together the problems of physical science.

Throughout the book, the *problems* play an important part. Many details of subject matter are referred to them, and some are designed to carry the student as far as his resources and imagination will permit. A few problems have no cut-and-dried solution. It is suggested that the majority of problems be either assigned or discussed in class.

The *references* to further reading have been chosen with the thought of avoiding those that are either too technical or not generally accessible; the instructor will find further bibliographies to historic works and critical essays in good histories of science in the several "Source Books" published by McGraw-Hill Co., in such well-known texts as that of Millikan, Roller, and Watson, and in that of L. W. Taylor, to whose bold experiment this book is considerably indebted. Bibliographies to the more modern papers will be found in the texts by Richtmyer and Kennard, Saha and Saha, Stranathan, and Cork.

I wish to acknowledge with pleasure my great debt to friends and colleagues whose advice aided in the preparation of this work—principally to Professors E. C. Kemble, P. LeCorbeiller, P. Frank, L. K. Nash, I. B. Cohen; Drs. T. S. Kuhn and C. L. Clark; Mr. D. H. D. Roller; to Professor F. W. Sears of M.I.T.; and to those indispensable critics and companions, my teaching assistants, particularly Mr. S. J. Smith, who prepared the major part of the problem answers.

To Professor Kemble goes an extraordinary measure of gratitude. In nine years of association, often teaching side-by-side with him in the same physics or general education course, I have not only learned from him, I fear I may unwittingly have appropriated from him whatever is of merit in this account.

*Part A*

THE STUDY OF MOTION

- 1 Velocity and Acceleration
- 2 Galileo and the Kinematics of Free Fall
- 3 Projectile Motion



*Galileo Galilei*



## THE STUDY OF MOTION

Historically as well as logically mechanics represents the foundation of physics and the prototype for the study of other physical sciences. The concepts developed in this field will appear again and again in this book. Moreover, mechanics is to physics what the skeleton is to the human figure—at first glance it may appear stiff, cold, and somewhat ghastly, but after even a brief study of its functions one experiences with mounting excitement the discovery of an astonishingly successful design, of a structure that is ingeniously complex, yet so simple as to be almost inevitable.

We shall begin with the key topic in mechanics, the laws governing some of the simpler motions of objects—movement along a straight line and a curved trajectory. For even in Galileo's time there was this axiom: *ignorato motu, ignoratur natura*.

## CHAPTER 1

### VELOCITY AND ACCELERATION

**1-1 Constant velocity.** This chapter will establish the concepts of velocity and acceleration, and although dealing with fairly simple material, it will also introduce some useful methods and cultivate some indispensable intuitions.

The very simplest of all problems of physics is a description of the *motion of a single object along a straight line*. Let us assume that we have been watching the motion of a car traveling along a perfectly straight and smooth road, and that we wish to find the exact relationship between the distance covered and the time required to cover it.

The progress of this car may be timed at stations regularly spaced at, say, 30 meters all along the line. The car has reached full speed before it goes past the first station, and all our time measurements refer to the moment that the first station is reached, the moment at which our "experiment" begins. If, by chance, the motion has been uniform, a tabular arrangement of the time schedule for the first five stations might read as in Table 1-1.

A more useful way of representing the same information is a time *vs.* distance graph. Along the horizontal axis (the *abscissa*) we plot

Table 1-1\*

STATION	Distance from station A in meters	Time of travel from station A in seconds
A	0	0
B	30	2
C	60	4
D	90	6
E	120	8

\* Whenever one takes experimental data such as appear in this table one must include a statement of the probable error of each type of measurement. Thus the second line in this table might read: B,  $30 \pm 0.1$ ,  $2 \pm 0.2$ . The *probable error* (as in the datum  $2 \pm 0.2$ ) is a technical term which signifies that the chances are fifty-fifty that the next measurement will fall between  $(2 + 0.2)$  meters and  $(2 - 0.2)$  meters. Evidently the determination of the probable error from a large mass of data is a technical and sometimes difficult matter.

the time elapsed in seconds, and along the vertical axis (the *ordinate*) the distance covered in meters\* (Fig. 1-1). On this graph the five known data are entered, and since they fall on a straight line, we draw it in as shown to indicate our not unreasonable belief that any additional data taken would also have fallen along that line. For example, we are fairly confident that at a point 45 meters from station A the car would have passed by 3 seconds after the beginning of our series of observations. This conclusion is an *interpolation* and is most conveniently done on a graph of this type. On the other hand, we are also fairly sure that after 9 seconds the vehicle, if it continued to move, reached a position 135 meters beyond station A. This is an *extrapolation*. Obviously, one must not place too much confidence in a truth obtained by interpolation, particularly when the basic data are few or widely spaced.\*\*

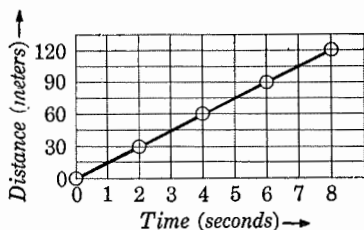


FIG. 1-1. Graph for uniform speed.

All this points to the generalization that for the particular motion which we here observed the distance covered is directly proportional to the time elapsed. Let the time elapsed since station A was passed be represented by the letter  $t$ , and let the distance covered during that time interval be called  $s$ ; then for the motion here described we may write the proportion

$$s \propto t.$$

A proportion can always be converted to an equation by introducing a constant which depends on neither of the variables. Therefore,

$$s = t \times \text{a constant};$$

\* Common units of measurement in scientific work are: for *length*, the meter (m), the centimeter (cm), and sometimes the millimeter (mm), where 1 m = 100 cm = 1000 mm = 39.37 in., or a little over 1 yard; for *time*, usually the second (sec); for *mass*, the kilogram (kgm) or the gram (gm), where 1 kgm = 1000 gm  $\doteq$  2.2 lb.

\*\* Note well several purely technical characteristics of this graph: (a) all known points, including the first one, are entered as points surrounded by a small circle; (b) the line is neat and is not drawn through the circles (why?); (c) the coordinate axes are accurately labeled, units are given. Reference to the original data should always be made. Follow these three rules in all plotting of graphs from experimental data, except when there are many experimental points close together; it may then be wiser to enter them as X.

consequently, 
$$\frac{s}{t} = \text{a constant.}$$

This ratio of distance covered to time elapsed is, by definition, the *velocity* or *speed* (we shall at present make no distinction between the two words), and it follows that our car travels with a constant or uniform speed.

Consider for a moment what an amazing thing has actually happened. First we watched the actual motion of a car along a straight road. Then, from the multitude of ever-changing impressions—the blur, the noise, the turning of wheels, the whole chaos of events progressing in time and space—we have rescued two measurable quantities,  $s$  and  $t$ , both of which take on different values every instant, and we have found that their ratio is a constant, an unchanging theme underlying the flux of otherwise meaningless, unrelated data. We have defined a concept, speed, and so have been led to discover a simple feature in an otherwise complex situation. Perhaps familiarity with the concept of *speed* prevents you from appreciating this experience of *creating order* from a *chaos* of sense impressions by abstracting from it some measurable data and by perceiving or “inventing” or intuiting a suitable concept to describe that portion of the total phenomenon; however, we shall return to this method, the very heart of scientific procedure, again and again.

The exact numerical value of the speed is to be found by substituting corresponding values for  $s$  and  $t$ . For example, from the data taken at station B,  $s = 30$  m and  $t = 2$  sec; from this, it follows that the speed,  $s/t$ , is 15 m/sec, which is about 44 ft/sec or 30 miles per hour (mi/hr).\* As a check on the constancy of the speed, you may wish to recompute it, using data for stations C, D, or E.

We have solved the problem concerning the motion of the car. The speed is constant, and is 15 m/sec. These two conclusions must seem obvious, but there are a few details which may be worth an extra word. To compute the speed we do not really need to rely on the raw data in Table 1-1. We can use the time *vs.* distance graph directly if we have any confidence in it at all. Consider, for example, the uniform motion indicated by the line  $OQPU$  in Fig. 1-2. From the previous paragraphs we know that for such a case the speed is given by  $s_1/t_1$  or by  $s_2/t_2$  (since  $s_1/t_1 = s_2/t_2$  by similar triangles), but it is equally well given by  $(s_2 - s_1)/(t_2 - t_1)$ , i.e., by the ratio of

---

\* The relationship  $30 \text{ mi/hr} \doteq 44 \text{ ft/sec} \doteq 13 \text{ m/sec}$  is a very useful one to memorize.

any small distance along the line of motion to the time needed to pass through that distance. This statement may seem intuitively acceptable, but we should check it rigorously.

PROBLEM 1-1. Prove that  $(s_2 - s_1)/(t_2 - t_1) = s_1/t_1$ .

SOLUTION A: Note that in Fig. 1-2 the distance  $QR = t_2 - t_1$ , and  $PR = s_2 - s_1$ ; also that the following three are similar triangles:  $Os_1Q$ ,  $Os_2P$ , and  $QRP$ . Therefore corresponding distances have equal ratios, that is,

$$\frac{s_1}{t_1} = \frac{s_2}{t_2} = \frac{s_2 - s_1}{t_2 - t_1} \quad \text{Q.E.D.}$$

SOLUTION B: Given that

$$\frac{s_1}{t_1} = \frac{s_2}{t_2},$$

prove that

$$\frac{s_2 - s_1}{t_2 - t_1} = \frac{s_1}{t_1}.$$

From the premise, it follows that

$$s_2 = \frac{s_1 t_2}{t_1},$$

therefore

$$s_2 - s_1 = \frac{s_1 t_2}{t_1} - s_1,$$

and

$$\frac{s_2 - s_1}{t_2 - t_1} = \left( \frac{s_1 t_2}{t_1} - s_1 \right) \left( \frac{1}{t_2 - t_1} \right),$$

which in turn reduces to

$$s_1 \left( \frac{t_2 - t_1}{t_1} \right) \left( \frac{1}{t_2 - t_1} \right),$$

that is

$$\frac{s_1}{t_1} \quad \text{Q.E.D.}$$

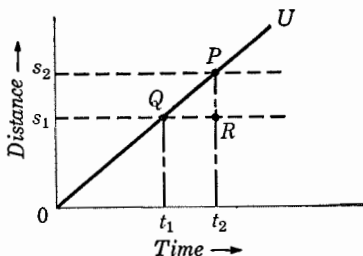


FIG. 1-2.

Before you leave these relatively harmless findings, be sure that you can answer to your own full satisfaction the following questions:

What is the definition of constant speed (use algebraic notation)?

What experiment must I make to find out whether a certain body moves with constant speed?

How do I find the value of the speed once I know that it is constant?

How small a distance ( $s_2 - s_1$ ) may I choose in order to calculate the speed?

After that, try to tackle the superficially deceptive questions upon which, as we shall find, depends the understanding of a large section of our physics:

For what reason do we here accept these arguments from Euclidean geometry or algebra?

What does mathematics tell us about real physical events such as the motion of material bodies?

Another problem is raised by a second critical look at our concept of constant speed along a straight line. Table 1-1 presented raw experimental data in tabular form and was our first step in formulating the idea of constant speed. Figure 1-1 represented these data in the form of a graph and expanded their usefulness by permitting convenient interpolation and extrapolation. The proportion  $s \propto t$  abstracted from this graph a concise statement which we expressed by the equation  $s/t = \text{a constant}$  (15 m/sec in this specific case).

What has happened here is that in four stages we have proceeded from a bulky table of raw data to a generalization ( $s/t = 15$  m/sec). If on repeated observations of this car we always arrive at this same equation, in other words, if the car always moves at a constant speed of 15 m/sec, we would regard the equation as a *law* applicable to this restricted type of observation. As it is, our equation arose from a single experiment and is simply a general description of this one motion.

At first glance all four stages (the table, the graph, the proportion, and the equation) seem to contain the same information, with the last stage appearing vastly preferable because of its more economical demands on paper and on mental effort. However, we must not let our innate love for conciseness misguide us; the equation  $s/t = 15$  m/sec does in fact include everything that the table offers, but it also tells indiscriminately a great deal more that may actually be quite unwarranted. For example, unless you had seen the tabulated data, you might believe on the basis of the equation alone that surely  $s = 1.5$  mm if  $t = 0.001$  sec, or that  $s = 4.7 \times 10^8$  m if  $t = 1$  year.

Both statements are correct only if the motion under investigation is truly uniform within the limits of such very exact measurements as 0.001 sec and for such a span of time as 1 year, but in our actual, limited experience neither of these conditions is likely to hold. The lesson to be drawn here is that equations in physics, other than

a few purely axiomatic definitions, must be thought to have attached an unseen "text," a statement which describes the actual limitations and other implied assumptions under which the equation may be applied. These conditions, this text, may simply be of the following form. "The equation  $s/t = 15$  m/sec describes the result of a certain experiment in which  $t$  did not exceed 8 seconds, and in which only a few measurements were made, each with such and such accuracy." Without the clear understanding that equations in physics almost always carry hidden limitations, we cannot expect to understand a single physical law; we would make unwarranted extrapolations and interpolations. We would be in the catastrophic position of a navigator who has to negotiate a rocky channel without having any idea of the length, width, and draft of his ship.

**1-2 The concept of average velocity.** Prolonged motion with constant velocity occurs so rarely in nature and is so limited a concept that, as you may have noticed, we have denied it the honor of a special letter symbol of its own. We turn now to motions with *nonuniform* velocity (accelerated or decelerated motions) but still taking place only along a straight line. A general case of this type is shown in Fig. 1-3, where a car starts from rest, gains speed until time  $t_1$ , and proceeds with constant speed until time  $t_2$  (as indicated by the straight-line portion between  $t_1$  and  $t_2$ ), after which it slows down to rest ( $t_3$ ), waits without moving until time  $t_4$ , and then again accelerates to reach a new, higher constant velocity. To deal with such variations in speed, it is necessary to introduce the familiar term, *average speed* or average velocity, symbolized by  $\bar{v}$ .

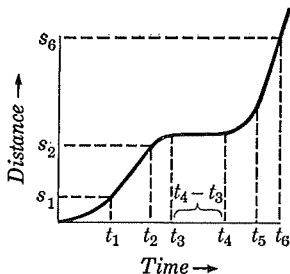


FIG. 1-3. Graph for nonuniform speed.

**DEFINITION:** *The average speed  $\bar{v}$  of an object during a time interval  $t$  is the distance covered during this time interval divided by  $t$ .*

Note that for any motion other than prolonged constant speed, the value of  $\bar{v}$  depends on the length of the time interval chosen. For example, a glance at Fig. 1-3 suffices to prove that  $\bar{v}$  for the first  $t_1$  seconds is smaller than  $\bar{v}$  computed for the first  $t_2$  seconds. On the other hand, during the interval  $(t_4 - t_3)$  the average speed is

zero. Although the speed from  $t_1$  to  $t_2$  might be 30 mi/hr and between  $t_5$  and  $t_6$  50 mi/hr, the average speed  $\bar{v}$  for the entire time to  $t_6$ , which is calculated by  $\bar{v} = (s_6/t_6)$ , is perhaps 20 mi/hr. We recognize that the meaning of  $\bar{v}$  is this: no matter how irregular the motion actually is during a time interval  $t$ , if the object had traveled instead with a constant speed of magnitude  $\bar{v}$  it would have traveled the same distance during those  $t$  seconds.

**1-3 Instantaneous velocity.** All our velocity measurements so far have involved more or less extended time measurements. However, we certainly also want to have a concept which tells us the velocity of a moving object *at one instant*, the type of information which we read off a speedometer. For example, we may ask, "What is the speed at time  $t_a$  for the motion pictured in Fig. 1-4?" Now we *do* know how to find the average velocity  $\bar{v}$  during a time interval from  $t_1$  to  $t_2$ , which includes the instant  $t_a$ , i.e.,  $\bar{v} = (s_2 - s_1)/(t_2 - t_1)$ . However, since the speed is not constant from  $t_1$  to  $t_2$ , we certainly cannot identify the velocity *at the instant*  $t_a$  with the average velocity  $\bar{v}$  *during*  $(t_2 - t_1)$ . Therefore we must resort to what undoubtedly will look like a trick: we calculate  $\bar{v}$  for a very short time interval  $\Delta t$  encompassing the instant  $t_a$ ; so short an interval that the value of  $\bar{v}$  would not change materially if the interval were made even shorter.\*

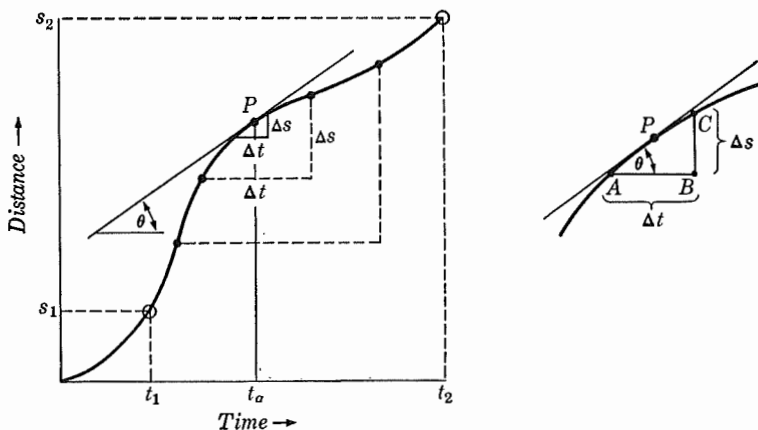


FIG. 1-4. Determination of instantaneous velocity.

\* The Greek letter  $\Delta$  (delta) will be used to indicate "a small change in . . ." Thus  $\Delta s$  does not mean "delta times  $s$ ," but reads "delta  $s$ ," i.e., a small increment in distance, a small displacement covered during the corresponding short time interval  $\Delta t$ . Therefore we can write  $\bar{v} = \Delta s/\Delta t$ .



The instantaneous velocity (symbolized simply by  $v$ ) at  $t_a$  is therefore calculated by successive approximations from the average velocity  $\bar{v}$ , and may be defined by

$$v = \lim_{\Delta t \rightarrow 0} \left( \frac{\Delta s}{\Delta t} \right).$$

Putting this equation into words, the instantaneous velocity is the value which the average velocity ( $\Delta s/\Delta t$ ) attains when the time interval ( $\Delta t$ ) is chosen sufficiently small, approaching zero in the limiting case.

A concrete example will help here. From the data on which Fig. 1-4 is based we might construct Table 1-2 for space and time intervals including point  $P$  on the curve. We note that for time intervals even shorter than the last one, 0.01 sec, we expect that  $\bar{v}$  will be only very little different from 1.7 cm/sec. Therefore, we conclude that the instantaneous speed  $v$  at time  $t_a$  is a little below 1.7 cm/sec.

Table 1-2

$t_2 - t_1 (\equiv \Delta t)$	$s_2 - s_1 (\equiv \Delta s)$	$\bar{v} = \frac{s_2 - s_1}{t_2 - t_1} = \frac{\Delta s}{\Delta t}$
10.0 sec	25.0 cm	2.5 cm/sec
1.0 sec	2.0 cm	2.0 cm/sec
0.10 sec	0.18 cm	1.8 cm/sec
0.010 sec	0.017 cm	1.7 cm/sec

As you perhaps remember from mathematics courses, what we have just done is the algebraic equivalent of finding the slope of the straight line tangent to the curve at the point  $P$  in Fig. 1-4. As  $\Delta t$  and  $\Delta s$  become smaller and smaller, the section of the curve which they include is more nearly a straight line, playing the role of the hypotenuse in the right triangle  $ABC$  (see insert in Fig. 1-4). But by trigonometry,  $\Delta s/\Delta t = \tan \theta$ , where  $\theta$  is the angle made by the hypotenuse with the horizontal; so that ultimately, when  $\Delta t$  is sufficiently small,  $\Delta s/\Delta t$  becomes simultaneously both the numerical value of the instantaneous speed  $v$  at point  $P$  and the value of  $\tan \theta$  of the line tangent at point  $P$ , i.e., the slope of that line.

We gratefully accept this unexpected result because it cuts short the labor involved in finding instantaneous velocities for nonuniform motions. In the future we need not construct such tables as Table

1-2 in order to find the instantaneous velocity at a given point during the motion of an object. We now simply refer to the distance *vs.* time graph, find the line that is tangent to the curve at the point in question, and calculate its slope. In fact, this last full sentence may stand as the definition of the concept of instantaneous speed in terms of actual operations needed for its determination. If ascertaining velocities by looking at a speedometer seems far simpler, realize that in the initial construction and calibration of the speedometer we must still take good care that it can automatically perform tasks equivalent to all those outlined in our definition of instantaneous speed.

It is this concept which we shall find most useful from now on, and we shall discuss typical problems in which the (instantaneous) velocity of a body changes from some initial value  $v_0$  to some other value  $v$  during a specified time interval  $t$ . We shall sometimes make plots of (instantaneous) velocity *vs.* time. For the case shown in Fig. 1-1, such a plot would evidently be simply a straight horizontal line, but Fig. 1-3 is more complex.

**PROBLEM 1-2.** Copy Fig. 1-3 on a large scale and superpose on it a plot of the velocity *vs.* time.

**1-4 Acceleration.** Of all motions, the one most important for our purposes is motion with *uniformly changing velocity*, i.e., motion with constant acceleration or constant deceleration. To a fair approximation this is the characteristic behavior of ordinary freely falling bodies, objects sliding down a smooth inclined plane, carts coasting to a stop on level ground, and generally of all bodies which experience a constant force.

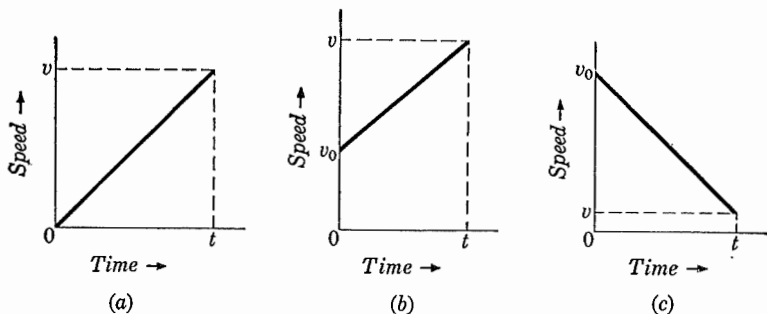


FIG. 1-5. Three cases of uniformly changing velocity.

Graphically, there are three simple cases involving uniformly changing velocity. In Fig. 1-5(a) the speed at the start is zero ( $v_0 = 0$ ) and reaches a value  $v$  after  $t$  sec; in Fig. 1-5(b)  $v_0$  is not zero, and the speed increases uniformly; and in Fig. 1-5(c) the speed decreases from a large initial value  $v_0$  to nearly zero during time  $t$ . Concrete experimental situations corresponding to each drawing will suggest themselves readily.

For all such cases it is easy to define the concept of acceleration, for which we shall use the letter  $a$ : acceleration is the ratio of the change of velocity ( $v - v_0$ ) to the time ( $t$ ) during which this change occurs, or

$$a = \frac{v - v_0}{t}. \quad (1-1)$$

As we have agreed to restrict our attention for the time being to motion with constant acceleration, we need not be concerned with such refinements as average acceleration and instantaneous acceleration, for, as in the case of constant velocity, the average and instantaneous values are numerically equal.

To illustrate this new concept let us consider some very straightforward cases.

*Example 1.* A skier starts down an incline and gathers speed with uniform acceleration, reaching a velocity of 30 ft/sec after 5 sec. What is his acceleration?

*Solution:* As in every problem, we must first *translate* each separate phrase into mathematical symbols, or else extract from each some clues as to which physical concepts, laws, and limitations are here applicable. The phrase "a skier *starts* down an incline" is translated to  $v_0 = 0$ ; "gathering speed with uniform acceleration" means  $a$  is constant, therefore Eq. (1-1) will hold for this case; "reaches a velocity of 30 ft/sec after 5 sec" can be simply restated as  $v = 30$  ft/sec,  $t = 5$  sec. The final phrase is, of course,  $a = ?$  Next we put all these findings side by side and read:

$$v_0 = 0; v = 30 \text{ ft/sec}; t = 5 \text{ sec}; a = \frac{v - v_0}{t} \text{ holds here}; a = ?$$

Evidently,  $a = \frac{30 \text{ ft/sec} - 0}{5 \text{ sec}} = 6 \text{ (ft/sec)/sec}$  (6 feet per second, per second, which may be written 6 ft/sec<sup>2</sup>).

Although this example must still appear exceedingly simple, you will find that the steps taken are the same here as in every single problem, no matter how sophisticated. If you train yourself from

the very start to *translate each phrase separately*, half the initial troubles with physics problems will have been avoided.

*Example 2.* A car has been traveling on a level road at a speed of 50 mi/hr when suddenly the motor fails and the car coasts along. During the first minute it loses 20 mi/hr. What was the acceleration during that time? (Evidently it will be a *negative* acceleration, called deceleration, but that is a superfluous term.)

*Solution:* Translating, we read  $v_0 = 50$  mi/hr,  $v = 30$  mi/hr,  $t = 1$  min = 60 sec,  $a = ?$  It is very tempting to plunge ahead and say that  $a = \frac{30 \text{ mi/hr} - 50 \text{ mi/hr}}{60 \text{ sec}} = -\frac{1}{3}$  (mi/hr)/sec. However, this is true only if we assume that the car coasted with uniform acceleration, a good guess but not necessarily the only correct one. Therefore we must say  $a = -\frac{1}{3}$  (mi/hr)/sec *if* the acceleration was uniform.

*Example 3.* A car on a roller coaster with many curves changed its speed from 5 mi/hr to 25 mi/hr in 20 seconds. What was the acceleration? Translation:  $v_0 = 5$  mi/hr,  $v = 25$  mi/hr,  $t = 20$  sec,  $a = ?$  However, since there is little likelihood that the acceleration was constant, we must simply confess that we cannot solve this problem until further data are given from which the type of motion may be deduced.\*

The units for acceleration were here variously given as (ft/sec)/sec or (mi/hr)/sec; others are evidently also possible, e.g., (m/sec)/min. In all these cases, the units correspond to the dimension of (length/time)/time, or in a symbolic shorthand frequently encountered,  $L/T^2$ . Here, as always, it is wise but not absolutely essential to convert all data to the same units of measurement (centimeters and seconds, or feet and seconds, etc.) before starting on the calculations.

**1-5 Equations of motion for constant acceleration.** The reason for our present preoccupation with uniformly accelerated motion is not only its occasional appearance in nature but also, and primarily, the *ease* with which we can formulate some simple equations of motions, i.e., equations which relate the five important variables: dis-

---

\* It would be a mistake to regard such a statement as an admission of dishonorable defeat; on the contrary, in science, as in every other field, the possibility of solving a problem offhand is the exception rather than the rule, and it is a good enough preliminary victory to know what additional information is needed for an eventual solution.

tance covered ( $s$ ), time taken ( $t$ ), initial velocity ( $v_0$ ), final velocity ( $v$ ), and acceleration ( $a$ ). Equation 1-1 is a good example. We may rewrite it as

$$v = v_0 + at \quad (1-2)$$

or, in words, the velocity at the end of a time interval  $t$  is the sum of the initial velocity and the change in velocity, which is the constant acceleration times the time elapsed.

There are other similar equations relating the five variables. Let us now construct three of them. An interesting and perhaps surprising conclusion implied in the speed  $vs.$  time graphs will help us: the area under the curve is numerically equal to the distance ( $s$ ) covered. For example, we redraw the second of the previous set of three graphs, as shown in Fig. 1-6. The area under the curve in Fig. 1-6(a) is now shaded, and may be thought of as a rectangle with sides  $v_0$  and  $t$ , surmounted by a triangle with height  $(v - v_0)$  and base  $t$ . The total area is therefore  $v_0t + \frac{1}{2}(v - v_0)t$ , which we can interpret as the distance  $v_0t$  covered during time  $t$  if the motion had proceeded with constant speed  $v_0$ , plus the distance  $\frac{1}{2}(v - v_0)t$  added by virtue of the acceleration of the motion. Our first conclusion is that we may write for the total distance covered

$$s = v_0t + \frac{1}{2}(v - v_0)t.$$

If we multiply and divide the second term by  $t$ , we obtain for it  $\frac{1}{2}\left(\frac{v - v_0}{t}\right)t^2$ , or, by Eq. (1-1),  $\frac{1}{2}at^2$ . Therefore,

$$s = v_0t + \frac{1}{2}at^2. \quad (1-3)$$

This is another of our four equations of motion.

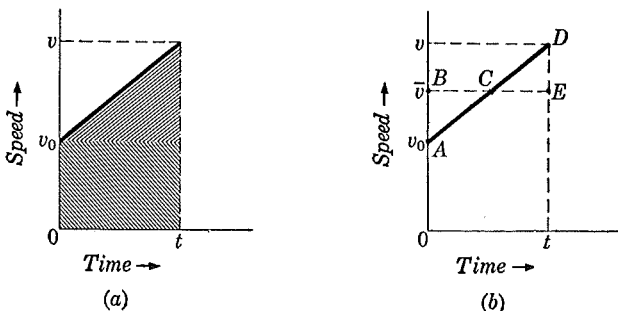


FIG. 1-6.

Our second conclusion is that the very same distance would have been covered if, instead of the accelerated motion pictured in Fig. 1-6, some constant average speed  $\bar{v}$  had been employed. By previous definition,  $\bar{v} = s/t$ , which on substitution becomes

$$\bar{v} = \frac{v_0 t + \frac{1}{2}[(v - v_0)/t]t^2}{t} = \frac{v_0 + v}{2}.$$

We rewrite this  $s/t = (v_0 + v)/2$ , or

$$s = \left(\frac{v_0 + v}{2}\right)t. \quad (1-4)$$

The statement that  $\bar{v} = (v_0 + v)/2$ , true of course only for motion with constant acceleration, has an intuitive appeal, as well as a graphical meaning which will be evident from Fig. 1-6(b). There, a horizontal line has been entered for  $\bar{v}$  at a level corresponding to  $(v_0 + v)/2$ . It is seen that the area under that line, and therefore the distance covered *at that average velocity*, is indeed equal to the distance covered with the *changing* speed during the same time interval  $t$ ; the whole proof depends, of course, on the equality of the two triangles  $ABC$  and  $CDE$ .

Finally, a fourth equation of motion may be derived by combining  $s/t = (v_0 + v)/2$  and  $a = (v - v_0)/t$  to give

$$v^2 = v_0^2 + 2as. \quad (1-5)$$

(Derive as a brief exercise.)

The last equation, though it is derived from the others and therefore can tell nothing new, still has a certain value which asserts itself more clearly when all four equations are juxtaposed. The following conventional grouping will show that we have here four combinations of the five variables  $s$ ,  $t$ ,  $v_0$ ,  $v$ , and  $a$ , allowing for the fact that only four of them are used at one time.

$$s = \left(\frac{v_0 + v}{2}\right)t \quad (\text{I})$$

$$s = v_0 t + \frac{1}{2}at^2 \quad (\text{II})$$

$$v = v_0 + at \quad (\text{III})$$

$$v^2 = v_0^2 + 2as \quad (\text{IV})$$

Now it is possible to step back, so to speak, and to study the over-all pattern of our work so far. We are now convinced that the equations I through IV contain all the important and interesting relations between distances traveled, time needed, and so forth, for

those motions which specifically are uniformly accelerated. Whenever we encounter this type of motion, be it executed by a falling stone or whatever, we shall at once know how to predict any two of the five variables when given the remaining three. For example, if it is known that a stone dropped from the Leaning Tower of Pisa strikes the ground 180 ft below with a measured speed of 107 ft/sec, we may at once calculate its time of fall and its acceleration. Translating the data to  $v_0 = 0$ ,  $v = 107$  ft/sec,  $s = 180$  ft,  $t = ?$ ,  $a = ?$ , we scan our four trusted equations and note that  $t$  can be found from equation I [ $t = (2s)/(v_0 + v) = 3.36$  sec] and the acceleration can be found either from IV [ $a = (v^2 - v_0^2)/2s$ ] or, now that  $t$  is known, from II or III.\*

The fact is that equations I through IV become our *tools* for dealing with this type of problem, and now that we have derived and established them we shall simply remember and apply them as occasion arises, in much the same manner as a craftsman, once having carefully selected a proper tool kit for his work, ceases to reinvent or justify his tools at every turn but is content to reach into his bag for the proper instrument demanded by the situation at hand. Before our tool kit begins to fill up, we must take the important step of gaining complete mastery over and confidence in each new concept or law as it is introduced. We can then apply it when called upon without feeling a need to re-examine it in detail. For example, when later we are to find  $v$  from  $v_0$ ,  $a$ , and  $s$  for a uniformly accelerating body, we can use equation IV (i.e.,  $v^2 = v_0^2 + 2as$ , therefore  $v = \sqrt{v_0^2 + 2as}$ ) even though undoubtedly one can have only little intuitive feeling about this relation. If you have initially derived equation IV to your satisfaction, you must then believe in it.

This is, of course, not at all the same thing as the self-defeating practice of just learning equations by heart or of looking them up in the notes to apply them when called upon. Such a ruse cannot work, because one cannot apply an equation properly without knowing all its intimate strengths and weaknesses, the implied "text." To give a final example on this belabored but important point, let us imagine that we are asked to find the time of fall of a brick from the top of the Empire State Building. Given are  $v_0$ ,  $a$ , and  $s$ . The

---

\* The assumption that  $a$  is indeed constant for free fall of real bodies at Pisa must be checked first if reason for doubt exists. If the motion should in fact turn out not to be uniformly accelerated, we should investigate how good an approximation our idealizations make toward the true case of free fall. This will be discussed in the next chapter.

temptation to "plug in" the readily memorized equation II must here yield to the knowledge which springs from closer study that the four equations undoubtedly do not apply in this case, since for such large distance of fall the air friction will prevent constancy of acceleration. As everyone knows by observation of falling leaves, snow, rain, etc., each body sooner or later may reach a constant terminal velocity of fall. This velocity, beyond which there is no further acceleration at all, depends on the surface area of the body, its weight, and the properties of the medium (air, etc.) through which it falls.

This completes our formal study of simple motion, the basis for our coming discussion of the laws of projectile motion.

### Additional Problems

NOTE: As a rule, do not spend more than about  $\frac{1}{2}$  hour on any problem at one session. If unable to solve by then, plan to come back to it later, or else finish it by stating what you foresee is likely to be the subsequent procedure and final solution. Note also that not all problems *have* simple numerical answers, even if the statement of the problem is numerical.

PROBLEM 1-3. (a) Find by graphical interpolation the approximate position of the car described in Fig. 1-1 corresponding to a time of travel of 5 sec, 0.5 sec, 7.005 sec. Extrapolate to find the position corresponding to 10 sec, 50 sec,  $-2$  sec (do this work on graph paper).

(b) Discuss the limitations of the process of interpolation and of extrapolation.

PROBLEM 1-4. On graph paper, plot a graph corresponding to Fig. 1-1, but use data which were taken at stations A, B, C, D, and E when a particle with a constant velocity of 50 mi/hr was passing by. You must first construct a table corresponding to Table 1-1. (Keep your eyes open; the numerous calculations can be simplified.)

PROBLEM 1-5. Describe the possible experimental situations which gave rise to each of the eight graphs of distance *vs.* time (Fig. 1-7) for motion along one line. Then try to express each of four or five of these graphs in terms of an equation relating *s* and *t* in algebraic fashion.

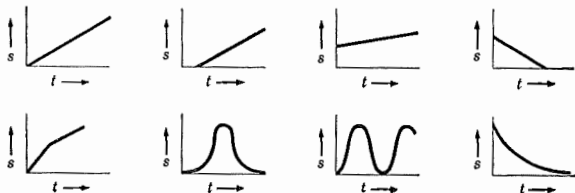


FIG. 1-7.



PROBLEM 1-6. A leisurely vacation trip from Cambridge, Mass., to Berkeley, Calif., took 15 days, including several long stops en route. Calculate the average speed for this trip in miles per hour. (NOTE: Whenever the data given seem insufficient, you will find the necessary additional information in pertinent books in your library. If you cannot find such books, use an intelligent guess.)

PROBLEM 1-7. Make some reasonable assumption about the speed which a car can pick up during the first 10 seconds after starting from rest, then calculate the acceleration.

PROBLEM 1-8. Prove first by algebra and then by geometric reasoning from Fig. 1-6(b) that  $\bar{v} = (v_0 + v)/2$  if the acceleration is uniform.

NOTE: In all future problems, assume that the acceleration of freely falling bodies at the surface of the earth is 32.2 (ft/sec)/sec or 980 cm/sec<sup>2</sup>. This happens to be true, within fairly narrow limits, over the whole earth. The symbol employed for this "constant" (the acceleration due to gravity) is the letter  $g$  instead of  $a$ .

PROBLEM 1-9. Five observation posts are spaced 20 meters apart along a straight vertical line. A freely falling body is released at the topmost station and is timed as it passes each of the others. Compute a table of the data expected, and represent this information in a graph (use graph paper).

PROBLEM 1-10. Make a table of the value of  $s$ ,  $v$ , and  $\bar{v}$  for a freely falling body if  $v_0 = 0$  and  $t$  is, in turn, 1 sec, 2 sec, 3 sec, 5 sec, and 10 sec. Represent this information in three graphs (on graph paper).

PROBLEM 1-11. A boy violently throws a stone vertically downward from a building 100 ft high, imparting to it a substantial initial speed. If the stone takes 1.8 sec to hit the pavement below, what was its initial speed?

PROBLEM 1-12. Make a list of all the relevant limitations, conditions, and idealizations which were implied in your solution of Problem 1-11.

PROBLEM 1-13. The nucleus of a helium atom is to travel along the inside of a straight hollow tube 2 meters long, which forms a portion of a particle accelerator. If the nucleus is to enter at 1000 m/sec and to leave at the other end at 10,000 m/sec, how long is it in the tube? State your assumption.

### Further Reading

R. A. Millikan, D. Roller, and E. C. Watson, *Mechanics, Molecular Physics, Heat, and Sound*. Boston: Ginn and Co. (1937). A successful, advanced text with frequent historical references. See pp. 7-22 for treatment of linear motion; pp. 457-458 for significant figures and notation by Powers of Ten.

## CHAPTER 2

### GALILEO AND THE KINEMATICS OF FREE FALL

**2-1 Introduction.** The material in the preceding chapter on velocity and acceleration for simple types of motion, which we have referred to as the logical and historical groundwork of physics, is essentially the work of Galileo Galilei. It is a small but critical part of the voluminous work of this arresting man, whom we must call the father of modern science. Now we shall examine in some detail Galileo's representative contribution to the problem of freely falling bodies as a historical, scientific, and methodological achievement. Although this field may seem to us fairly self-evident, it is actually a very challenging one. One of the foremost contemporary historians has said, "Of all the intellectual hurdles which the human mind has been faced with and has overcome in the last fifteen hundred years the one which seems to me to have been the most amazing in character and the most stupendous in the scope of its consequences is the one relating to the problem of motion . . ."

Galileo is important and interesting to us for several reasons. First, if we understand what he had to struggle against, what contemporary attitudes he had to discard in attaching himself to the "New Philosophy," then perhaps we shall be clearer about the nature of our present-day structure of science and our own attitudes and preconceptions. To cite an analogy, we certainly must not fail to study the period of the American Revolution if we wish to understand our present democratic institutions, to which it gave rise.

Second, after the historical setting of Galileo's work has been clarified, he shall speak to us as a scientist at his best, as an investigator whose virtuosity in argument and discovery excites in his listeners a deep and lasting experience. He shall be for us an example of a physicist at work, and although in his case circumstances were extraordinary at almost every turn, so far is he from being unrepresentative that if he were to come to life again now, he very probably would, after a period of study, continue his work most excellently in many branches of science in his own characteristic manner.

Third, Galileo's approach to the problem of motion of freely falling bodies will provide us, both in this chapter and later, with an opportunity for a short introspective discussion on procedure in science. We need this. Without an awareness of the place of defini-

tion, hypothesis, experimentation, and intuition in physical law and physical research, without a knowledge of the vital processes by which this field functions and grows, science must seem a dry formalism to be memorized without understanding and reproduced without enthusiasm.

To begin with, let us consider what kinematics\* Galileo might have learned at his university, what type of physics he eventually had to reject. It was, of course, the science of the scholastics, the heirs and often misinterpreters of that greatest of scientist-philosophers of antiquity, Aristotle. But at this point, a warning: the history of science is an erudite and difficult field in its own right. By the very nature of its subject matter, there is a good deal of inconclusiveness and uncertainty at important points, and we shall in this book neither attempt nor presume to evaluate such delicate questions as, for example, to what extent Galileo shared many opinions and methods with the Aristotelians, or drew on the work of other investigators—Archimedes, the whole tradition of scholars at the venerable University of Paris, Tartaglia, Benedetti, or Stevinus.\*\* For science, like any other intellectual enterprise, cannot start from scratch without any roots in some tradition. Galileo, too, had the often unacknowledged encouragement of many enlightened bits of contemporary physics in his struggle to shape a consistent whole. But his fight, judging by his own words, does appear to have been against a group of exponents of the stricter contemporary scholastic view. In short, the statement that Galileo started modern science is so abrupt and leaves out so much that it can justly be called wrong, but no more wrong than the statement that the American Republic dates from June, 1788, when the Constitution went into effect.

**2-2 Aristotelian physics.** Despite the widespread story of Galileo's experiment from the Leaning Tower of Pisa, most people believe (erroneously) that heavy bodies gain speed far more quickly when falling than do lighter bodies. This was the view recorded by Aristotle (384-322 B.C.), Plato's great pupil, much of whose works, through the mediation of St. Thomas Aquinas and his followers, became the basis of Christian thinking in Europe from the thirteenth century on. However, Aristotle's view on the motion of falling bodies

---

\* Kinematics = physics of moving bodies, excluding the study of the specific action of forces.

\*\* For details on the state of mechanics before Galileo, particularly on the motion of projectiles, see Chapter IV in Henry Crew's *The Rise of Modern Physics*. Baltimore: Williams and Wilkins, 1935.

was not simply an error or the result of miscarried experimentation, but a characteristic consequence of a great over-all scheme, a very minor member of an imposing structure which Aristotle erected to search for a unity of *all* thought and experience through conceptual schemes, embracing at the same time elements that we now separate into component parts: scientific, poetic, theological, ethical.

Proceeding from a then current notion, Aristotle postulated that all actual matter within our physical reach is a mixture of *four* elements, Earth, Water, Air, and Fire, it being understood that a lump of actual earth or a stone might contain much of the singly unobtainable pure element Earth in the presence of a certain admixture of the other three elements. Thus a vessel full of the purest water can be expected to contain beside the element Water also some earthy substance; and indeed, on boiling away the liquid, one can find a solid residue. So convincing was this experiment and this interpretation that when the great chemist Lavoisier by ingenious and exact experimental demonstration first proved in 1770 that the residue came from the vessel itself, he also reported that what was "thought by the old philosophers, still is thought by some chemists of the day."

A second postulate prescribed that each of these four elements carried a tendency or desire to reach its "natural place" of repose, which for Air and Fire is above our earth, and for Water and Earth is downward as seen by us. A third postulate directed that the actual motion of an object is determined by the tendency of the element most abundantly present. So the behavior of steam rising from a boiling vessel was explained as the upward motion owing to the introduction of the element Fire into the heated water; when, on cooling, the steam gave up its Fire, the then predominant element, Water, could assert itself again, and the condensed moisture could precipitate down to its natural place below.

A partial consequence of this view was that the motion of an object up or down to its "natural place," the so-called "natural motion," was so governed by the balance of inherent elements that its speed of motion must be proportional to the amount of the predominant element. A big stone, containing evidently more Earth than a small one, consequently was expected to descend that much faster when allowed to execute its "natural motion" of free fall. And in truth, as everyone knows, generally the heavy bodies do outstrip lighter ones, particularly if we compare extremes.

We are, of course, quick to enter the objection that as a matter of experimental fact heavy and light stones do not at all show the

ratios of speeds predicted by Aristotle, that, in fact, they would fall *equally* fast if it were not for the often negligible effect of air resistance, and that therefore the logically complete chain of Aristotelian reasoning breaks down at the starting point, namely at the failure of the postulates. Now this type of objection at once shows up our modern criteria of scientific truth, namely, the need that all our theorizing must proceed from some observable facts and in the end again must yield accurate and preferably quantitative predictions of other observable facts. Not so the disciples of Aristotle. As he might put it, what can be observed is not the final touchstone of the theory, but rather a more or less valid illustration thereof. True, the theory itself springs in part from experienced fact—Aristotle in particular held that knowledge proceeds from experience, and he was a gifted observer, as his work in biology shows—but his theory must do more than describe and predict facts; it must give them meaning in a wider sense, by showing that the facts are in accord with the general postulates of the whole philosophical system. To the Aristotelian of the late 16th century the postulates concerning elements and tendencies, and all conclusions from them, were true (“philosophically true”), clear, certain, because they were also part of a larger, important, and satisfying scheme, reaching into theology, which could not be given up just because the detailed motion of two weights dropped from a tower contradicted one of its many conclusions. He would give “philosophic truth” precedence over contradictory “scientific truth.” In short, though he may seem to deal with the motion of bodies and other problems of interest to a modern scientist, the Aristotelian was not and did not intend to be a scientist as we conceive of it. His physics was the investigation of the nature of things in a context so large that the preoccupation of modern science would seem to him like a small, artificial, incomplete subject with little significance of its own. As F. S. Taylor puts it in *Science Past and Present*:

“The science of Aristotle and of the scholastics had been primarily a science of purpose. It was transparently clear to the latter that the world and all that’s in it was created for the service of man, and that man had been created for the service of God. That was a perfectly intelligible scheme of the world. The sun was there to give us light and to tell the time and mark out the calendar by his motions. The stars and planets were a means of distributing beneficent or maleficent influences to the things on earth to which they were sympathetically linked; plants and ani-

mals were there to give us food and pleasure, and we were there to please God by doing His will. Modern science does not produce any evidence that all this is not perfectly true, but it does not regard it as a scientific explanation; it is not what science wants to know about the world."

Conversely, in general (though there are notable exceptions), students of nature before Galileo's time would necessarily have found it unreasonable or even inconceivable to be seriously interested only in the type of questions our science asks. That such questions are indeed interesting or important, that the methods employed in our scientific investigations can give results at all, this could not have been apparent before the results themselves began to pour forth in the 17th century. But so overwhelming was the unsuspected, ever-growing stream that we, who are caught in the flood, can hardly think in prescientific terms. We must deplore this unresourceful imagination of ours. If each of us even for a moment could transcend his own pattern to appreciate clearly the point of view from which an Aristotelian would have proceeded, we should be better equipped for dealing with the everyday conflicts of differing temperaments and opinions.

We must not forget that the Aristotelian's view of nature offered him two further features: none of his conclusions and predictions were quantitatively so rigorous or so absurd that he could not overlook the quantitative disagreement; it is only the skeptic who will go to the trouble of climbing high towers and of burdening himself with cumbersome weights for a precise test—and even he will probably find the results far from clear-cut in his first few trials. Second, by endowing Nature with tendencies and purposes in her every particle, the Aristotelian made her understandable in terms of his own human longings and reactions; she was then explainable by intelligible purposes, which was to him far more satisfying emotionally than the new scientific world of descriptive "cold" facts could ever be.\* As has recently been said by the eminent French historian of science, Alexandre Koyré:

"Therefore what the founders of modern science, among them Galileo, had to do, was not to criticize and to combat certain faulty

---

\* To avoid the danger of regarding this summary as a general pronouncement on Greek or even Aristotelian science, read by all means Benjamin Farrington's *Greek Science* (Pelican Books A142 and A192); note particularly the account of ingenious though largely qualitative experimental work attributed to several different schools of thought.

theories, and to correct or to replace them by better ones. They had to do something quite different. They had to destroy one world and to replace it by another. They had to reshape the framework of our intellect itself, to restate and to reform its concepts, to evolve a new approach to Being, a new concept of knowledge, a new concept of science—and even to replace a pretty natural approach, that of common sense, by another which is not natural at all.”

To be more specific about Aristotle's teachings on the mechanics of motion—and here he almost certainly gave voice not to a new discovery but to a current notion—we must rely on a few lines out of the whole monumental bulk of his work, some passages from his books *De Caelo* and his *Physica*, which state that the time of free fall in a given medium is inversely proportional to the weight of the object. These details of natural, vertically downward motion for heavy matter (or upward motion for “light” matter like smoke or steam) are, however, not at all applicable to “celestial” objects, which rather consist of a specific fifth element (the quintessence, aether), eternal as contrasted with the four changeable terrestrial elements, and endowed with a natural *circular* motion as compared with terrestrial straight-line motion. Furthermore, when a terrestrial object is propelled in a manner contrary to its natural motion (e.g., a stone or other projectile being thrown upward), it will exhibit “violent” or “forced” motion for a time until the effect wears off and natural motion may again take place. A typical 16th century view of the motion of a projectile, directly influenced by Aristotle, is sketched in Fig. 2-1. The long-observed increase of velocity of a *falling* stone was sometimes explained as the manifestation of the increasing purity or desire of the element Earth in the stone to reach its natural place, just as a traveler approaching his home after an absence is possessed by deepening longings and so hastens his steps.

Several further conclusions follow from this scheme, for example, that even for a supposedly uniform and unaccelerated motion along the first part of the path a force on the projectile was considered essential to the maintenance of motion, this force being supplied by the vortices of air

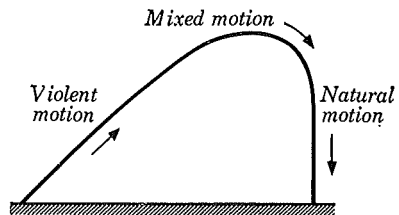


FIG. 2-1. An early sixteenth-century conception of projectile motion.

around the projectile. This particular point, the common-sense view that a net force is required to maintain a uniform velocity, was one of the serious stumbling blocks in the way of physics which helped to arouse the ire of generations of commentators like Sir Edmund Whittaker, who has written recently that Aristotle's science "was worthless and misleading from beginning to end," and Bertrand Russell, who thought that Aristotle "has been one of the great misfortunes of the human race."

**2-3 Galileo. Biographical sketch.** Galileo Galilei, our symbol of the great struggle out of which science arose, was born in Pisa in 1564, the year of Shakespeare's birth and Michelangelo's death. His father was a poor but noble and spirited Florentine from whom he acquired an active and competent interest in poetry, music, and the classics. Galileo's ingenious inventiveness began to show itself early, when as a medical student at the University of Pisa he invented a simple pendulum-type timing device for the accurate measurement of pulse rates. It appears that even at that time he made himself conspicuous by constantly challenging the opinion of authority.

Lured from medicine to physical sciences by the reading of Euclid and Archimedes, he quickly became known for his unusual powers, and at 26 he was appointed to the University Chair of Mathematics. His characteristic traits now showed themselves: an independence of spirit and an inquisitive and forthright intellect unmellowed by tact or patience. Soon after his appointment he seems to have begun to attack his colleagues, almost without exception, of course, persuaded Aristotelians and, like disciples in general, rabid defenders of their often erroneous interpretations of the master's works. This is the time when he may have made a public experiment on the equal speed of unequal weights dropped from the famous Campanile of Pisa, though it is quite likely that the story is only a legend. Because the incident is so important historically, so widely known, and so frequently debated, we may, for what it is worth, look at the source of the story—the biographical notes on Galileo by one of his latest and closest students, Vincenzo Viviani—an account which has often been called vague and somewhat romantic.

"As it seemed to him [Galileo] that a true knowledge of the nature of motion was required for the investigation of the natural effects, he gave himself completely to the contemplation of that [motion]: and then to the great confusion of all the philosophers, very many conclusions of Aristotle himself about the nature of



motion, which had been theretofore held as most clear and indubitable, were convicted of falseness, by means of experiments and by sound demonstrations and discourses; as, among others, that the velocity of moving bodies of same composition, unequal in weight, moving through the same medium, do not attain the proportion of their weights, as Aristotle assigned it to them, but rather that they move with equal velocity, proving this by repeated experiments performed from the summit of the Campanile of Pisa, in the presence of all other teachers and philosophers and of all students."\*

Truth or fable, Galileo must have known and would not have failed to point out that experiments with such results had been made and reported on previously. He was beginning to gain enemies by his blunt attacks against the Aristotelians for their obedience to the word of ancient authority.

Before Galileo's tenure was completed he had to leave Pisa, evidently forced out by his enraged enemies. Later we find him at Padua in the less doctrine-conscious republic of Venice. Here he begins his crucial work on astronomy on behalf of the Copernican theory, which later brought him further enemies and great suffering as well as immortal fame.

Drawn back to his native Tuscany in 1610 by a generous offer of the Grand Duke, Galileo became his Court Mathematician and Philosopher—a title he chose himself. From then until his death at 78 in 1642, his life, despite recurring illness and family and money troubles, was filled with continuous excellent work, with teaching and writing, and of course with the relentless and tragic fight with his enemies.\*\* He left us an orientation toward the world of phenomena which demanded an explanation in terms of simple, all-pervading natural law and so helped to sweep out the host of angels, intelligences, ultimate causes, and animistic devices which governed the world picture of most of his contemporaries. Above all he represents the mathematical approach to nature, a philosophy that has its roots in the Pythagorean and Platonic tradition, and so found itself in conflict with the qualitative approach to nature that marked the Aristotelian tradition.

Among Galileo's specific achievements we find essentially the first

---

\* For this new translation I am indebted to Dr. Arnolfo Ferruolo of Harvard University.

\*\* For a biography of Galileo, read J. J. Fahie's *Galileo, His Life and Work*. New York: James Pott, 1903 (still very useful).

integrated advances in physics since the time of Archimedes over 1800 years before: the foundation of our mechanics, and his arguments on behalf of the Copernican theory of planetary motion, so influential in the fight to establish the heliocentric theory. This connection will lead us back to Galileo as a key personality in a later chapter.

Among his many inventions we count a thermometer, a telescope, an improved microscope, the design of the first pendulum clock—symbols of Galileo's love for and preoccupation with actual mechanical devices and the problems of craftsmen, perhaps of just those in his own instrument shop, where he personally made instruments including, during a single year, hundreds of telescope lenses. This contact with actual experimental problems may have stimulated his remarkable intuitive appreciation of natural law, which helped him even when the mathematical argument now and then became dubious, as must have happened frequently in his days when algebraic notation, decimal notation, logarithms, analytical geometry, etc., were unknown or just being accepted, and when some of the most necessary mathematical methods of present-day physics were not yet in sight. In fact, it is this very point which partly explains why the work of Galileo later had to be supplemented, transformed, and given its full meaning by Isaac Newton, born within the year after Galileo's death. As has been said with much insight, Galileo represented the assault and Newton the victory.

**2-4 Galileo's Two New Sciences.** Of interest to us now is Galileo's book *Discourses and Mathematical Demonstrations Concerning Two New Sciences Pertaining to Mechanics and Local Motion*. It was written while Galileo was a technical prisoner of the Inquisition, and was published surreptitiously in Holland in 1638, for in his own Italy his books were then being suppressed.

Galileo, even though he was then old, sick, and going blind, writes in an enjoyable style, as we have seen.\* He introduces three men who discuss the manuscript of a fourth, of Galileo himself. Of the speakers, *Simplicio* represents the Aristotelian view; he is attacked by *Salviati*, while the third, *Sagredo*, is pictured as an informed, helpful friend of the others. Of the several parts of the book, the

---

\* Translation after that of Crew and De Salvio, but incorporating a few changes. The italics are not in the original. A much condensed series of excerpts is in J. W. Knedler's *Masterworks of Science*, Garden City: Doubleday and Co., 1947. This latter work is to be used with great caution, however, because it is not made clear in the text where the original has been abbreviated.

third and fourth contain the material on "local motion," i.e., motion with uniform velocity and with uniform acceleration, and projectile motion. Let us allow ourselves a rare luxury and listen to them as they discuss freely falling bodies:

"*Salviati*: I greatly doubt that Aristotle ever tested by experiment whether it be true that two stones, one weighing ten times as much as the other, if allowed to fall, at the same instant, from a height of, say, 100 cubits,\* would so differ in speed that when the heavier had reached the ground, the other would not have fallen more than 10 cubits . . .

"*Sagredo*: . . . I, who have made the test, can assure you that a cannon ball weighing one or two hundred pounds, or even more, will not reach the ground by as much as a span ahead of a musket ball weighing only half a pound . . .

"*Simplicio*: Your discussion is really admirable; yet I do not find it easy to believe that a bird shot falls as swiftly as a cannon ball.

"*Salviati*: Why not say a grain of sand as rapidly as a grindstone? But, *Simplicio*, I trust you will not follow the example of many others who divert the discussion from its main intent and fasten upon some statement of mine which lacks a hairbreadth of the truth and, under this hair, hide the fault of another which is as big as a ship's cable. Aristotle says that an iron ball of one hundred pounds falling from a height of one hundred cubits reaches the ground before a one-pound ball has fallen a single cubit. I say that they arrive at the same time. You find, on making the experiment, that the larger outstrips the smaller by two finger breadths . . . ; now you would not hide behind these two fingers the ninety-nine cubits of Aristotle, nor would you mention my small error and at the same time pass over in silence his very large one."

This is as clear a statement as we would wish of the idea that a naive first glance at natural events is not at all a sufficient basis for a physical theory; different freely falling bodies in air indeed do *not* arrive at the very same instant at the foot of the tower, but this, on further thought, turns out to be far less significant than the fact that they do *almost* arrive at the same instant. Attention to this last fact, a fact which is actually far removed from the immediate sense

---

\* One cubit is about 18-20 inches; 100 cubits happens to be nearly the height of the Tower of Pisa.

impressions gained in this experiment, is suggestive and fruitful, because it regards the failure of equal arrivals as a minor error instead of a major truth, as an experimental circumstance explainable by air friction. And indeed, when the air pump was invented soon after the time of Galileo, this hypothesis was confirmed by the observation of the truly simultaneous fall of a light feather and a heavy coin inside an evacuated glass tube. We are reminded of the old quip that "science has grown almost more by what it has learned to ignore than by what it has had to take into account." Here again we meet the recurring theme in experimental science: regard observable events with a penetrating eye to search behind the immediate confusion of appearances for an underlying simplicity and mathematical lawfulness. In this particular case, everything depended on being able to "think away" the air and its effect on free fall. This is easy enough for us who know of pumps and the law of gases; but it was at that time conceptually almost impossible for several reasons, some evident, but some quite subtle. For example, Aristotle, who had examined the question of the existence of a vacuum, had denied its possibility as contrary not only to experience but to reason. He argued, where there is a vacuum there is *nothing*. If nothing remains, he continued, then also space itself and the physical laws can no longer exist in this vacuum; for how could a stone, for example, know which way is up or down except in relation to and through the influence of the surrounding atmosphere?

The answer depended on dissociating the words *vacuum* and *nothingness*, and preventing *space* from becoming annihilated merely because air has been dismissed or withdrawn. But though Galileo found it possible to do this mentally, we may believe the rest of science accepted the existence of the vacuum mainly on the basis of subsequent *experimental demonstrations* that falling bodies, heat, light, magnetism, etc., were indeed propagated through vessels from which air had been withdrawn.

Others had known before Galileo that the Aristotelians were wrong about free fall, but it is to his credit that he proceeded to discover the details of the correct description of this motion and to make it part of a more general system of mechanics. The task he set himself in the *Two New Sciences* was to invent concepts, methods of calculation and measurement, and so forth, to arrive at a description of the motion of objects in rigorous mathematical form. In the following excerpts we must not lose sight of the main plan. First, he will discuss the mathematics of a *possible*, simple type of motion, defined

as uniformly accelerated. Then he will assume or hypothesize that this is the very motion of actual free fall under usual conditions. Third, he will derive experimentally testable conclusions from his hypothesis. Last, he will by experiment confirm the assumption, proving thereby the constancy of acceleration in free fall. And so he introduces the discussion with these remarks:

“My purpose is to set forth a very new science dealing with a very ancient subject. There is, in nature, perhaps nothing older than motion, concerning which the books written by philosophers are neither few nor small; nevertheless I have discovered some properties of it that are worth knowing and that have not hitherto been either observed or demonstrated. Some superficial observations have been made, as, for instance, that the free motion of a heavy falling body is continuously accelerated; but to just what extent this acceleration occurs has not yet been announced; . . .

“It has been observed that missiles and projectiles describe a curved path of some sort; however no one has pointed out the fact that this path is a parabola. But this and other facts, not few in number or less worth knowing, I have succeeded in proving; *and what I consider more important*, there have been opened up to this vast and most excellent science, of which my work is merely the beginning, ways and means by which other minds more acute than mine will explore its remote corners.

“This discussion is divided into three parts: the first part deals with motion that is steady, or uniform; the second part, with motion as we find it accelerated in nature; the third part, with so-called violent motions and with projectiles.”

There follows first a very thorough and searching discussion of uniform (unaccelerated) motion, along the lines indicated in our previous chapter, and then the section on “naturally accelerated motion”:

“We shall consider now naturally accelerated motion, such as that of falling bodies.

“First of all it seems desirable to find and explain a definition that best fits natural phenomena. Although certainly it is quite permissible to invent an arbitrary type of motion and study its course (as, indeed, some have proposed to themselves helices and conchoids as the paths described in certain motions—it being granted that nature does not make use of such—and have very commendably deduced the properties of these from their definitions), we however have decided to consider the phenomena of

freely falling heavy bodies with an acceleration such as actually occurs in nature, and to make our definition of accelerated motion exhibit the essential features of this type of natural accelerated motion. And this, at last, after repeated efforts we trust we have succeeded in doing. In this belief we are confirmed mainly by the consideration that experimental results are seen to agree with and exactly correspond with those properties which have been described by us.

“Finally, in the investigation of naturally accelerated motion we have been led, as if by the hand, by careful observation of the usage and precept of nature herself in all her other actions, in which she is accustomed to employ the simplest and easiest means . . .

“. . . When, therefore, I observe a stone initially at rest falling from an elevated position and continually acquiring new increments of velocity, why should I not believe that such increases take place in a manner which is exceedingly simple and fairly easily apprehended by everybody? If now we examine the matter carefully, we find no addition or increment more simple than that which repeats itself always in the same manner. This we readily understand when we consider the intimate relationship between time and motion; for just as uniform motion is defined by and thought of in terms of equal time intervals and equal distances (thus we call a motion uniform when equal distances are traversed during equal time intervals), so also we may, in similar manner, by thinking in terms of equal time intervals, conceive additions of velocity as taking place without complication . . . , hence the definition of the motion which we are about to discuss may be stated as follows:

*“A body is said to be uniformly accelerated when, starting from rest, it acquires equal increments of velocity during equal time intervals.*

*“Sagredo:* Although I can offer no rational objection to this or indeed to any other definition, devised by any author whosoever, since all definitions are arbitrary, I may nevertheless without offense be allowed to doubt whether such a definition as the foregoing, established in an abstract manner, corresponds to and describes that kind of accelerated motion which we meet in nature in the case of freely falling bodies . . . .”

The discussion now turns toward the only correct answer to this excellent question, namely, that Galileo's “arbitrary” definition of

acceleration (which we may write as  $a = (v - v_0)/t$  for this case of constant acceleration) happens to be most useful to describe the experimental facts of real, observable motion. But there is first a significant little excursion when Sagredo proposes:

“From these considerations perhaps we can obtain an answer to a question that has been argued by philosophers, namely, what is the *cause* of the acceleration of the natural motion of heavy bodies . . .”

Salviati sternly turns away this persistent preoccupation of the previous two millennia with a simple and modern type of statement which, roughly speaking, directs us to ask first about the “How” instead of the “Why” of motion, not to waste energy on a theory of motion while the *descriptive law* of motion is yet unknown:

“*Salviati*: The present does not seem to be the proper time to investigate the cause of the acceleration of natural motion concerning which various opinions have been expressed by various philosophers, some explaining it by attraction to the center, others to repulsion between the very small parts of the body, while still others attribute it to a certain stress in the surrounding medium which closes in behind the falling body and drives it from one of its positions to another. Now, all these fantasies, and others too, ought to be examined; but it is not really worth while. At present it is the purpose of our Author [Galileo] merely to investigate and to demonstrate some of the properties of accelerated motion (whatever the cause of this acceleration may be).”

Galileo then scrutinizes his decision that for uniformly accelerated motion he should regard the increase in velocity proportional to time elapsed rather than distance covered, a proposition which equally well fulfills his requirement of simplicity of relationships. He admits having at one time entertained the latter alternative, but shows by a set of plausible experiments which he asks us to imagine (in a set of *thought experiments*, one of the very fruitful devices in the scientist's kit of methods) that such a hypothetical motion does not correspond to actual free fall.\* In modern terms, a motion for which it is true that  $(v - v_0) \propto s$  may arbitrarily be *called* uniformly accelerated, but in this case the word acceleration cannot mean  $a = (v - v_0)/t$ , and cannot be the same constant which we have postulated to be the constant during actual free fall.

\* His argument at this point, though correct in its conclusion, is erroneous in some details.

Having disposed of this alternative, Galileo might now have produced a set of experimental data to show that in free fall, say for a stone dropped from a high tower, the measured quantity  $(v - v_0)/t$  does indeed remain constant for various time intervals  $t$  during the descent. This would have settled everything at once. But consider the experimental troubles involved in such a direct attack! Today we could conceivably use high-speed motion picture equipment to record the successive positions of the falling stone, from which a graph of distance *vs.* time could be conducted, which in turn should permit calculation of the necessary data for an instantaneous velocity *vs.* time graph and enable us to check the hypothesis that  $(v - v_0)/t = \text{constant}$  throughout the motion. But Galileo's age did not even have a good clock for timing rapid motions. Experiments on fast free fall being impossible, he now characteristically turned to a more easily testable consequence of his hypothesis, "the truth of which will be established when we find that the inferences from it correspond to and agree exactly with experiment."

First he convinced himself analytically that a ball rolling down a smooth incline obeys the same kind of rules which apply to motion in free fall, that it is in short a "diluted" or slowed-down case of that motion, whatever its law. If the ball is found to move with constant acceleration, so must also a freely falling body. His attention therefore turned finally to this simple experiment:

"We took a piece of wooden scantling, about 12 cubits long, half a cubit wide, and three fingerbreadths thick. In its top edge we cut a straight channel a little more than one finger in breadth; this groove was made smooth by lining it with parchment, polished as smooth as possible, to facilitate the rolling in it of a smooth and very round ball made of the hardest bronze. Having placed the scantling in a sloping position by raising one end some one or two cubits above the other, we let the ball roll down the channel, noting, in a manner presently to be described, the time required for the descent. We repeated this experiment more than once in order to be sure of the time of descent and found that the deviation between two observations never exceeded one-tenth of a pulse beat. Having performed this operation until assured of its reliability, we now let the ball roll down only one-quarter the length of the channel; and having measured the time of its descent, we found it to be precisely one-half of the former. Next we tried other distances, comparing the time for the whole length with that for the half, or for two-thirds, or for three-fourths, or indeed for



any fraction. In such experiments, repeated a full hundred times, we always found that the distances traversed were to each other as the squares of the times, and this was true for any inclination of the . . . channel along which we rolled the ball . . .

“For the measurement of time, we employed a large vessel of water placed in an elevated position; to the bottom of this vessel was soldered a pipe of small diameter giving a thin jet of water, which we collected in a small cup during the time of each descent, whether for the whole length of the channel or for a part of its length; the water thus collected was weighed on a very accurate balance; the differences and ratios of these weights gave us the differences and ratios of the time intervals, and this with such accuracy that, although the operation was repeated many, many times, there was no appreciable discrepancy in the results.”

Do not fail to notice the further ingenious twist in the argument. Even in these inclined plane experiments it was not possible to test directly whether  $(v - v_0)/t = \text{constant}$ , because this would have demanded direct measurements of instantaneous velocities  $(v, v_0)$ , whereas Galileo could measure directly only distances and time intervals. Therefore he decided to test whether another equation of motion from the previously discussed set holds; for example, the equation  $s = v_0t + \frac{1}{2}at^2$ , an equally good test of constant acceleration, particularly if  $v_0 = 0$  (ball started from rest) and therefore  $s \propto t^2$ . The actual experimental confirmation that the distance which the ball rolls down the incline was indeed proportional to the square of the elapsed time for a given angle of incline was both ample proof of the hypothesis and also a rather trustworthy conclusion, since it involved only the two measurable quantities  $s$  and  $t$ .

Finally, note also two aspects of Galileo's account, which, in contrast to some other features of his work, are so modern that you will find them in most reports of contemporary experimental research. He so fully describes his experiment in every essential detail that all qualified investigators can attempt to repeat his observations and to check on their validity. Second, he repeats his important experiments many times and under various circumstances, not only to become intimately acquainted with and stimulated by all features of the total experience, but also to estimate the error inherent in his measurements and to test the reliability of his work. Both of these aspects have ever since been among the most important rules of this fruitful game.

Although we have to leave this particular problem of motion for the time being, Galileo found it possible and enjoyable to enlarge on his findings in a detailed and ingenious manner, and to deduce and confirm other conclusions from his fundamental hypothesis of naturally accelerated motion. One conclusion which offered itself to him, and which was apparently then regarded as a fairly inconspicuous result, was first given precise form by Descartes in 1644 and, as it turned out, later became the key which in the hands of Newton was to unlock the floodgates of physical discovery. It is a crude form of what we now call the *Law of Inertia*:

“We may remark that any velocity once imparted to a moving body will be rigidly maintained as long as there are no causes of acceleration or retardation, a condition that is found only on horizontal planes [where the force of friction has also been minimized]; for in the case of planes that slope downwards there is already present a cause of acceleration, while on planes sloping upward there is retardation; from this it follows that motion along a horizontal plane is perpetual . . .”\*

In modern language we should say that *in the absence of a net force*, be it gravity, friction, or whatever, *a body will remain in unchanging motion or else at rest*. This amounts to a redefinition of the concept “force,” a negation of the Aristotelian view that a force is needed to maintain unchanging motion. Furthermore, it represents a profound change of attitude in theology and cosmology, for by implying that not only rest but also uniform motion is a “natural” state for ordinary terrestrial bodies, Galileo gave them an attribute previously reserved for celestial bodies only, boldly uniting terrestrial and celestial phenomena in *one* physics applicable to both.

This may at the moment seem like a minor matter, but we shall have to agree with the great mathematician and philosopher, A. N. Whitehead, who said of the law of inertia, “This is the first article of the creed of science . . . It is more than a mere statement of belief: it is a paean of triumph over defeated heretics. It should be set to music and chanted in the halls of Universities.”

Galileo himself was not unaware of the potentialities inherent in his pioneering work on motion. As he had introduced it, he also

\* Note here Galileo's frequently invoked *method of the limiting case*, which has ever since been one of the standard tools of scientific thought; he reasons about a *hypothetical* motion on a straight horizontal frictionless plane by considering it as the limiting case of *observable* motion on inclined planes.

concludes his treatment of naturally accelerated motion on a prophetic note:

“The theorems set forth in this brief discussion, if they come into the hands of other investigators, will continually lead to wonderful new knowledge. It is conceivable that in such a manner a worthy treatment may be gradually extended to all the realms of nature.”

### Problems

PROBLEM 2-1. What exactly are our present-day objections to Aristotle's system of elements? Why might these objections be unimportant in the eyes of an Aristotelian?

PROBLEM 2-2. Read through Galileo's "Third Day" in the *Discourses on Two New Sciences* in order to find out what his arguments were for thinking that the law of motions on an inclined plane are the same as for free fall. (For a translation use that of H. Crew and A. de Salvio, Macmillan, 1933, or the much condensed version in J. W. Knedler's book.)

PROBLEM 2-3. List the steps outlined in this chapter by which Galileo progressed from his first definition of what he means by uniformly accelerated motion to his final confirmation that his definition coincides with the actually observable motion of free fall. What limitations and idealizations entered into the argument?

PROBLEM 2-4. Read through Galileo's "Third Day" to Theorem IV. Then restate as concisely as possible the argument that the acceleration  $a$  of an object sliding on a smooth inclined plane is to the acceleration in free fall  $g$  as the vertical height of the plane  $h$  to the length of the plane  $l$ , or, in short, that  $a = g \sin \theta$  (Fig. 2-2).

PROBLEM 2-5. Summarize what specific methods of attack and of solution you may have discerned in this discussion of Galileo's work on freely falling bodies.

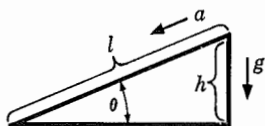


FIG. 2-2.

### Further Reading

*Note:* The books mentioned in these pages are intended to guide you to stimulating sources if you wish to carry this subject further. The excerpts from the original writings are especially recommended, for, as Mach wrote, "There is no grander nor more intellectually elevating spectacle than that of the utterances of the fundamental investigators in their gigantic power."

H. Butterfield, *The Origins of Modern Science*. London: G. Bell and Sons (1949), Chapters I and V. A very penetrating account.

H. Crew, *The Rise of Modern Physics*. Baltimore: Williams and Wilkins (1935), Chapter IV.

J. J. Fahie, *Galileo, His Life and Work*. New York: James Pott and Co. (1903).

Galileo Galilei, *Dialogues Concerning Two New Sciences*, transl. by H. Crew and A. De Salvio. New York: Dover Publications (1952). First and Third Days.

J. W. Knedler, Jr., *Masterworks of Science*. Garden City: Doubleday and Co. (1947). Biographical notes and condensed excerpts, e.g., pp. 75-151, Galileo.

W. F. Magie, *A Source Book in Physics*. New York: McGraw-Hill (1935). Short and valuable biographical notes and excerpts, e.g., Galileo's work on falling bodies, pp. 1-17.

A. N. Whitehead, *Science and The Modern World*. New York: Macmillan Co. (1925); also Pelican Mentor Book M28. Chapter I.

A. N. Whitehead, *Essays in Science and Philosophy*. New York: Philosophical Library (1947), pp. 227-242.

*Note:* The thought processes which may have guided Galileo in his development of the laws of motion are discussed from the point of view of Gestalt psychology with insight and lucidity in Max Wertheimer, *Productive Thinking*, New York: Harper & Bros. (1945).

## CHAPTER 3

### PROJECTILE MOTION

Turning now to the more general motion of projectiles, we leave the relatively simple case of movements along a *straight line* only and expand our methods to deal with motion *in a plane*. Our whole understanding of this field, a typical and historically important one, will hinge on a far-reaching discovery: The observed motion of a projectile may be thought of as the result of two *separate* motions, combined and followed *simultaneously* by the projectile, the one component of motion being an unchanging, unaccelerated horizontal translation, the other component being a vertical, accelerating motion obeying the laws of free fall. Furthermore, these two components do not impede or interfere with each other; on the contrary, the resultant at any moment is the simple effect of a superposition of the two individual components.

Again it was Galileo who perfected this subject, in the section "The Fourth Day" immediately following the excerpts from the *Two New Sciences* which were presented in our last chapter. Let us re-fashion his original arguments to yield somewhat more general conclusions, and to illustrate by one painstaking and searching inquiry how an understanding of a complex problem may grow. In later chapters it will usually be necessary that you supply some of the intermediate steps of a problem, therefore this effort now will be a good investment. Thus, if Chapter 2 represented a historically inclined chapter, this one may be called a factually directed one.

**3-1 Projectile with initial horizontal motion.** We may start with two simple but perhaps surprising experiments.

(a) If we watch a plane in steady horizontal flight drop a small, heavy parcel, we realize that (except for air friction) the parcel *remains directly below the plane* while, of course, dropping closer and closer to the ground. This is represented in Fig. 3-1. If we watched this event from a

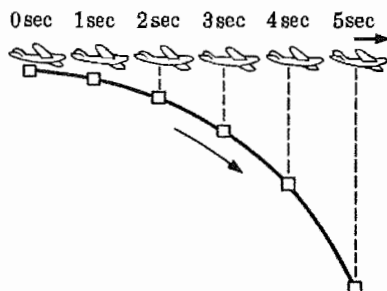


FIG. 3-1. Successive positions of a parcel dropped from a plane

H. Crew, *The Rise of Modern Physics*. Baltimore: Williams and Wilkins (1935), Chapter IV.

J. J. Fahie, *Galileo, His Life and Work*. New York: James Pott and Co. (1903).

Galileo Galilei, *Dialogues Concerning Two New Sciences*, transl. by H. Crew and A. De Salvio. New York: Dover Publications (1952). First and Third Days.

J. W. Knedler, Jr., *Masterworks of Science*. Garden City: Doubleday and Co. (1947). Biographical notes and condensed excerpts, e.g., pp. 75-151, Galileo.

W. F. Magie, *A Source Book in Physics*. New York: McGraw-Hill (1935). Short and valuable biographical notes and excerpts, e.g., Galileo's work on falling bodies, pp. 1-17.

A. N. Whitehead, *Science and The Modern World*. New York: Macmillan Co. (1925); also Pelican Mentor Book M28. Chapter I.

A. N. Whitehead, *Essays in Science and Philosophy*. New York: Philosophical Library (1947), pp. 227-242.

*Note:* The thought processes which may have guided Galileo in his development of the laws of motion are discussed from the point of view of Gestalt psychology with insight and lucidity in Max Wertheimer, *Productive Thinking*, New York: Harper & Bros. (1945).

balloon, from very high up and directly above this region, we should, of course, see only the horizontal part (component) of the motion; and if from there we now could see the parcel at all, we should think it traveled directly along with the plane, instead of falling away from it. The clear implication is that the horizontal component of the motion of the parcel remains what it was at the moment of release from the plane (unchanged, as suggested by the law of inertia), even though there is superposed on it the other, ever-increasing component of vertical velocity.

(b) In the second experiment we shall place two similar small spheres at the edge of a table (*cf.* Fig. 3-2). At the very moment that one is given a strong horizontal push, so that it flies off rapidly and makes a wide, curved trajectory,\* the other is touched only very gently and therefore drops straight down to the floor. Which will land first?

The experimental fact is that they both land together, no matter what the initial horizontal speed of the pushed ball; furthermore, the balls remain at equal levels throughout their fall, though of course they separate in the horizontal dimension, one having a large (and by the previous experiment, presumably undiminishing) horizontal velocity and the other having none. The conclusion here is that the vertical component of the motion is quite independent of any additional horizontal movement.

The joint result of both experiments is that motion in a plane can be resolved into two components, the horizontal and vertical, and in the case of projectiles (parcel, spheres, etc.) moving near the earth, these two components of motion are independent. This allows us to answer some inquiries into the motion of our projectiles, the simplest being "How far away from the starting point does the parcel or the second sphere land?" During the time taken for the fall or flight, say  $t$  seconds, the horizontal, unchanging component of motion will transport the projectile a distance equal to the initial velocity imparted in the forward direction at the moment of release multiplied by the time  $t$ ; simultaneously the growing vertical component of motion contributes a downward displacement by an amount corresponding to the distance of free fall during that time  $t$  under the influence

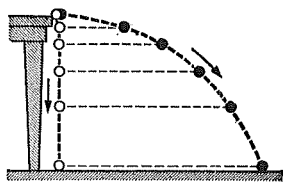


FIG. 3-2. The horizontal and vertical components of motion are independent.

\* By trajectory we mean here the path followed by a projectile under the influence of gravity.

of the gravitational acceleration,  $g$ . If we agree to give the subscript  $x$  to all horizontal components, and  $y$  to all vertical components, as indicated in Fig. 3-3, then we may rewrite the previous sentence more concisely:

$$s_x = v_{0x}t \quad \text{or} \quad s_x = v_x t, \quad (3-1)$$

because  $v_{0x} = v_x$ , the horizontal motion being by assumption unaccelerated; and secondly,

$$s_y = \frac{1}{2} g t^2, \quad (3-2)$$

because  $v_{0y}$  was here assumed to be zero, the initial motion at the time of release having been altogether in the horizontal direction.

If  $t$  is the actual total time of flight,  $s_x$  so calculated is the so-called "range" of the projectile and  $s_y$  is the distance of drop, but our question above inquired about neither  $s_x$  nor  $s_y$ . If we still wish to find the total displacement from the starting point (a quantity we might call  $s$ ), a glance at Fig. 3-3 will at once solve our problem. By the Pythagorean theorem, for displacement in a plane,

$$s = \sqrt{s_x^2 + s_y^2}. \quad (3-3)$$

We may now substitute, and find that  $s = \sqrt{(v_{0x}t)^2 + (\frac{1}{2} g t^2)^2}$ . In order to produce a numerical answer, we must therefore be given the initial velocity and time of flight, quite apart from the value of  $g$ , which in all our work will be assumed to be numerically equal to 980 cm/sec<sup>2</sup> or 32.2 ft/sec<sup>2</sup>.

Furthermore, the angle which direction  $s$  makes with the horizontal,  $\beta$ , is clearly given\* by  $\tan \beta = s_y/s_x$ . It is at this angle that a "sight" will have to be set to ascertain the proper instant of release of this projectile from the plane. It would be a more difficult problem, and one which we shall not try to solve, to discover not the total *displacement* but the length of the actual path along the curved trajectory. It happens to be a fairly unimportant question to us, the range  $s_x$  and the distance  $s_y$  being the more sought-after quantities; yet we may deduce an interesting result about the shape of the path. From Eqs. (3-1) and (3-2),

$$s_x = v_{0x}t, \quad \therefore t = \frac{s_x}{v_{0x}} \quad \text{and} \quad t^2 = \frac{s_x^2}{v_{0x}^2}.$$

\* See Appendix V for a summary of trigonometric relations.

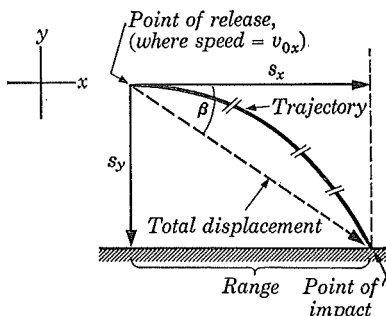


FIG. 3-3. Trajectory and total displacement of a projectile.



But

$$s_y = \frac{1}{2} g t^2 = \frac{1}{2} g \left( \frac{s_x^2}{v_{0x}^2} \right),$$

which may be written

$$s_y = \left( \frac{g}{2v_{0x}^2} \right) s_x^2. \quad (3-4)$$

Since the quantities in parentheses are constant throughout a particular projectile motion, it appears that  $s_y \propto s_x^2$ . This, however, in analytic notation, is the property of a simple parabola (Fig. 3-4).

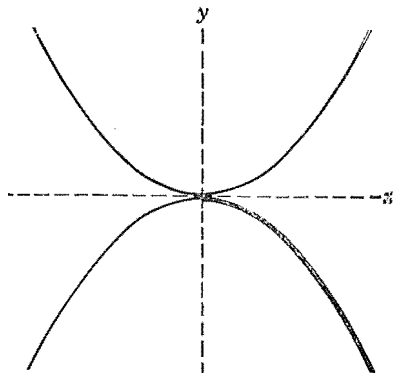


FIG. 3-4. Complete graph of the equation  $y = (\text{a constant}) \cdot x^2$ . For the present we shall refer only to the right limb of the lower parabola (heavy line).

**PROBLEM 3-1.** Plot on graph paper the equation  $y = (-\frac{1}{3})x^2$ . (a) Give  $x$  only + values; (b) give  $x$  both + and - values. Recall the useful convention that to distances above the origin or to the right of the origin, + signs are assigned; for distances below or to the left of the origin, - signs.

After Galileo had deduced the parabolic nature of the trajectory (by an original and similar argument), projectile motion immediately became much simpler to understand; for the geometrical properties of the parabola had long been established by the mathematicians in their disinterested pursuit of abstract geometrical worlds. We find here a first clue to three important facts of life in science: (1) If we can express phenomena quantitatively and cast the relation between observables into equation form, then we can grasp the phenomenon at one glance, manipulate it by the laws of mathematics, and so open the way to the discovery of new truths concerning this phenomenon. For example, having found that our trajectories are parabolic, we would confidently calculate, if required, the length of the actual path along the curve by means of some formula proposed to us by a mathematician, one who may never even have seen an actual projectile motion, but who has studied parabolas thoroughly. (2) Consequently there is always an imperative need for a well-developed system of pure mathematics from which the physicist may draw. (3) We can see why the physical scientist always tries to cast his problem into such a form that established approaches, procedures, or tricks from another branch of science or from mathematics will

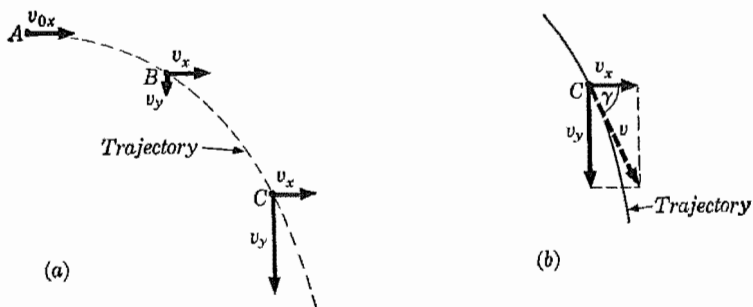


FIG. 3-5. The actual velocity  $v$  can be compounded from its components  $v_x$  and  $v_y$ .

aid in the solution. As an example of the last point, just as Galileo applied the mathematics of parabolas to actual projectile motions, so does the modern acoustical engineer solve his problems by means of the mathematical schemes developed quite independently by electrical engineers for their own very complex field. Whatever the methods of science may be, they have shown themselves to be transferable from one specialty to the other in a remarkable and fruitful way.

Yet another question to be asked about our projectile motion is this: What is the actual velocity of the moving object at some time  $t$  after release? We may be prompted by our initial two experiments and by the discussion leading to Eq. (3-3) to represent the two observed components of the actual velocity at several portions along the path, as has been done in Fig. 3-5(a). At each of the three selected points the horizontal velocity component  $v_{0x} = v_x = \text{constant}$ , but the vertical component grows linearly from  $v_{0y} = 0$  to a larger and larger value, in accordance with the law for falling bodies,  $v_y = v_{0y} + gt$ , which here becomes simply  $v_y = gt$ . Our experiments showed us that  $v_x$  and  $v_y$  will not disturb each other, but this does not tell us yet what is the actual velocity  $v$  at any chosen point on the trajectory. Now it was Galileo's very important discovery that the total effect of both velocity components, i.e., the actual velocity  $v$ , could be compounded from the two components  $v_x$  and  $v_y$  by the very same simple scheme which enabled us to find  $s$  when  $s_x$  and  $s_y$  were known. In Fig. 3-5(b), the velocity at point C is constructed by completing the parallelogram (in this case simply a rectangle, since  $v_x$  and  $v_y$  are at  $90^\circ$ ) formed by the arrows corresponding to  $v_x$  and  $v_y$ ; the total, the "real" velocity, is then indicated by the

diagonal  $v$ , which at every point is tangent to the momentary path of the projectile, and which has the value

$$v = \sqrt{v_x^2 + v_y^2}. \quad (3-5)$$

Consequently, the momentary direction of motion at point  $C$  can be specified in terms of the angle  $\gamma$  between  $v$  and the horizontal direction, and is given by

$$\tan \gamma = \frac{v_y}{v_x}. \quad (3-6)$$

This, to repeat, is a postulate that is *checked by experiment*. We may cast it into formal language: the instantaneous velocity of a projectile, and its direction of motion at any instant, may be obtained by *superposing the two independent components  $v_x$  and  $v_y$  in a simple parallelogram construction*, where  $v$ , the resultant, is given in both direction and magnitude by the diagonal, as shown in Fig. 3-5(b).

This *principle of the superposition of velocity components* is so simple as to seem now self-evident, but that is deceptive. The *only* justification for postulating the crucial equations (3-5) and (3-6) is that they are essentially experimentally confirmed discoveries. In one form or another the superposition principles are such simplifying devices that they may seem to have been too much to hope for; yet natural phenomena graciously do exhibit such features at times, and it is practically unthinkable that physics could have embarked on its great journey of vigorous discovery from Galileo on without his having intuited many such "simple" fundamentals.

*Example 1.* A stone is dropped from a plane flying level with the ground at a speed of 200 ft/sec. If the stone needs 5 seconds to reach the ground, how high did the plane fly, what is the range of this projectile, what is its total displacement, and what is its final speed just before striking? As we now translate phrase by phrase, we write:  $v_{0y} = 0$ ,  $v_{0x} = 200$  ft/sec,  $t = 5$  sec,  $s_y = ?$   $s_x = ?$   $s = ?$   $v = ?$  We must assume that  $g = 32.2$  ft/sec<sup>2</sup> (often, 32 ft/sec<sup>2</sup> will be accurate enough), and that air friction is so small on the projectile that its horizontal motion is not affected and its vertical motion remains uniformly accelerated. Both of these assumptions are, incidentally, experimentally found to hold quite well for the first few seconds of motion, but increasingly less well thereafter.

For a solution we find  $s_y = \frac{1}{2}gt^2 = 400$  ft;  $s_x = v_{0x}t = 1000$  ft,  $s = \sqrt{400^2 + 1000^2}$  ft, and so forth.

**3-2 Introduction to vectors.** The foregoing discussion has exhausted every worthwhile question we may put at present about projectile motion, *provided the motion started with a straight horizontal velocity* and no vertical component of velocity at all. Now we shall go on to the general case, i.e., to projectile motion with all conceivable initial conditions of velocity, not neglecting projectiles shot from a gun pointing at any angle above or below the horizontal line. Our interest here is not so much a practical one—those aspects of science which men abuse to direct weapons against one another are surely the most dismal. Instead, we shall regard this exposition as a typical formulation of a large and successful *general scheme to handle all possible specific instances* of a problem. Furthermore, in this lengthy effort we shall of necessity have to develop or invent some new physical concepts of permanent usefulness.

Imagine now the progress of a projectile which leaves the muzzle of a gun with an initial velocity  $v_0$  and at an angle  $\theta$  with the horizontal [Fig. 3-6(a)], rises to some height  $h$ , and then drops into a valley below the horizon. The horizontal and vertical components of the initial velocity,  $v_{0x}$  and  $v_{0y}$ , can be computed by *resolution* from  $v_0$  and the angle  $\theta$ , by means of the same trigonometric approach which led us to the converse proportion, the *composition* of  $v$  from  $v_x$  and  $v_y$ .

$$\begin{aligned} \text{Horizontal component of } v_0 &= v_{0x} = v_0 \cos \theta, \\ \text{Vertical component of } v_0 &= v_{0y} = v_0 \sin \theta. \end{aligned} \tag{3-7}$$

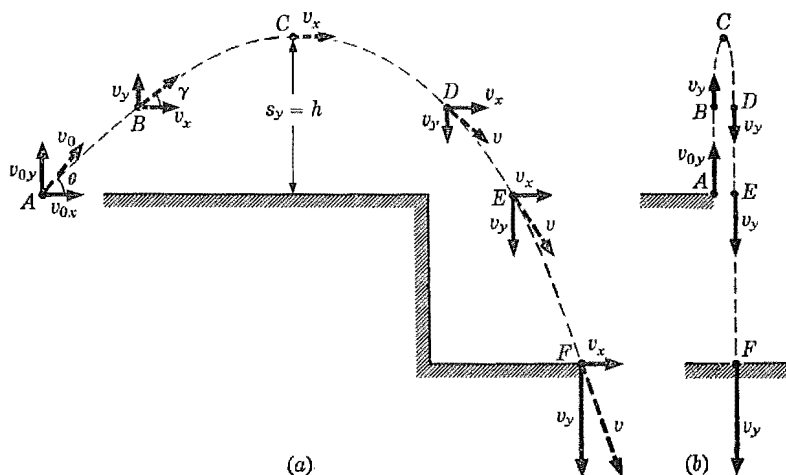


FIG. 3-6. Generalized projectile motion.

As before,  $v_{0x}$  remains also the value of the horizontal component  $v_x$  at all later points, including the six positions shown. ( $F$  refers to the state of affairs just *before* the projectile hits.) On the other hand,  $v_{0y}$  at point  $A$  is subsequently diminished as the projectile rises against the opposing, downward acceleration of gravity, until, at point  $C$ ,  $v_y = 0$ . From then on,  $v_y$  no longer points upward, but reverses and becomes larger and larger in the previously examined manner. *So far as this vertical component alone is concerned*, its history would have been exactly the same if  $v_{0x}$  had been zero, i.e., if we had been thinking about a stone thrown vertically upward at the edge of a cliff with that same initial value of  $v_{0y}$  [Fig. 3-6(b)].

The question now arises how to calculate the total actual velocity  $v$  of the projectile in Fig. 3-6(a) at some point  $B$ . We still have faith that at  $B$ , as elsewhere,  $v = \sqrt{v_x^2 + v_y^2}$  (and in the absence of faith an experiment would prove the point).  $v_x$  at  $B$  is still, in magnitude, equal to  $v_{0x}$ , and therefore equal to  $v_0 \cos \theta$ . But when we now turn to the vertical component, we find that its initial magnitude has been decreasing between points  $A$  and  $B$  through constant deceleration due to gravitational pull in a direction opposite to the motion. We may regard  $v_y$  at  $B$  as a constant speed component  $v_{0y}$  diminished by a contrary, growing component of magnitude  $gt$ , where  $t$  is the time needed to reach point  $B$  (see Fig. 3-7, "at  $B$ "). At point  $C$ , evidently  $v_y = 0$ , since there the term  $gt$  has become exactly as large as  $v_{0y}$ . Beyond that point we may become somewhat uncertain how to calculate  $v_y$  from  $v_{0y}$ ,  $t$ , and  $g$ , and here we must have recourse to a geometric scheme.

Consider Fig. 3-6(b). Each arrow representing  $v_y$  at any one point may itself be thought of as the result of simple addition or subtraction of the length of two other arrows, representing respectively the constant velocity  $v_{0y}$  (upward), and the cumulative effect of gravity,  $gt$ , downward (Fig. 3-7). But this is just a special case of a general scheme of additions of quantities such as displacement (for which we used it in Section 3-1) and velocities, the scheme due to Galileo which we previously employed to compose  $v$  from  $v_x$  and  $v_y$ : no matter whether two velocities simultaneously at work on one body are active at right angles to each other (as before),

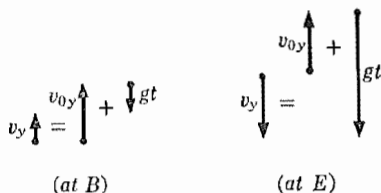


FIG. 3-7.

or along the same line (as now while we consider only the *vertical* component), or make any other angle with each other, we can find the resulting total velocity by the so-called parallelogram method of construction. In this construction, the arrows representing the two velocity components are drawn to some scale (e.g., 1 cm = 10 ft/sec), placed on paper in correct relation end to end, the parallelogram is completed, and the diagonal is drawn which, pointing in the direction away from the meeting point of the component arrows, gives to the same previous scale the magnitude and direction of the resultant total velocity.

In Fig. 3-8 several such hypothetical cases are shown,  $v$  being the resultant of two components  $v_a$  and  $v_b$ . The first of these examples is familiar to us. The last is a kind of degenerate case of the one preceding; the two components lie in opposite directions along one line and the drawing of a clear parallelogram becomes practically impossible, although the result is obviously that shown. Similarly, if the angle between  $v_a$  and  $v_b$  in the third sketch were to shrink to zero, the resultant,  $v$ , would grow to the plain numerical sum of the components. Quantities which permit such a graphic method of addition are called *vectors*; besides displacement and velocity we shall find many other important vector quantities, notably force. In each case the method for vector addition (that is, for finding the resultant of two separate components) is this parallelogram method, based on a simple superposition law.

**3-3 The general case of projectile motion.** We shall soon have ample opportunity to make use of vector addition in other contexts, but for the moment it is enough to consider the scheme in Fig. 3-7 as a restatement of the special case of vector addition shown in the last drawing of Fig. 3-8. Now a simple convention suggests itself which circumvents the unavoidably inaccurate and somewhat messy scheme of drawing these arrows to scale to find the resultant in each case. If we agree to the convention to give *positive* values to quantities represented by an upright arrow, and *negative* values to those oppositely directed, then we can employ the simple equation

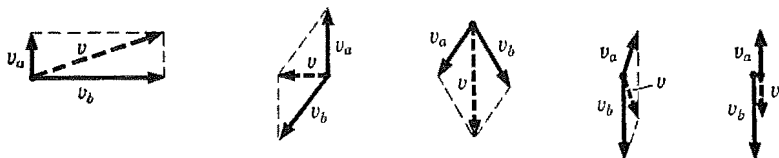


FIG. 3-8. Graphical method for adding two velocity vectors.

$$v_y = v_{0y} + gt \quad (3-8)$$

in its original form to compute  $v_y$  at any time  $t$  after the projectile motion starts. A glance at Fig. 3-7 will show qualitatively that this equation, together with that convention, assures the correct calculation of the numerical values for  $v$ .

*Example 1.* A projectile is shot from a gun tilted at  $60^\circ$  above the horizontal; the muzzle velocity  $v_0$  is 3000 ft/sec. (a) What is  $v_y$  after 20 sec of flight, (b) what is the total velocity  $v$  of the projectile after 20 sec of flight, (c) at what angle with the horizontal does it then travel?

*Solution:* (a)  $v_y = v_{0y} + gt$ , where  $v_{0y} = v_0 \sin \theta = 2600$  ft/sec (positive value);  $gt$  is numerically about  $32$  (ft/sec<sup>2</sup>)  $\times 20$  sec = 640 ft/sec, but representing always a "downward directed" quantity, it is by our convention to be used with a negative sign, that is, as  $(-640)$  ft/sec. Therefore  $v_y = 2600$  ft/sec +  $(-640)$  ft/sec = 1960 ft/sec. [The fact that our answer has a positive sign itself means that the velocity is still directed upward; this means we might be now at some position on the trajectory like point  $B$  in Fig. 3-6(a).] (b) To compute  $v$  by  $\sqrt{v_x^2 + v_y^2}$ , and then the angle from  $\tan \gamma = v_y/v_x$ , we now find  $v_x$  from  $v_x = v_{0x} = v_0 \cos \theta$ . (Continue as exercise.)

*Example 2.* How long will the above-mentioned projectile take to reach the highest point? The translation of this question depends on the realization that  $v_y = 0$  at the top point of the trajectory, i.e., at point  $C$  in Fig. 3-5(a). Then  $gt = -v_{0y}$  by Eq. (3-8), or  $(-32 \times t)$  ft/sec =  $-2600$  ft/sec, and consequently  $t = -2600/(-32)$  sec  $\doteq 81$  sec. (We enter at once our suspicion that for such conditions the physical realities of air friction undoubtedly make our calculations quite academic.)

*Example 3.* What is  $v_y$ , the vertical component of velocity of this projectile, after 162 seconds?

*Solution:* From Eq. (3-8) we find  $v_y = 2600$  ft/sec +  $(-32 \times 162)$  ft/sec  $\doteq 2600$  ft/sec - 5200 ft/sec =  $-2600$  ft/sec (negative = downward). In short, after a time as long again as it took to reach the top of the trajectory, the velocity component downward is equal numerically but oppositely directed to the initial vertical component. This is pictured at point  $E$ , Fig. 3-6.

The last two examples make it clear that it might be wise to assign the minus sign permanently to the value for  $g$ , to write  $g = -32$  ft/sec<sup>2</sup> once and for all, and so banish the temptation to change the plus sign in Eq. (3-8) at any time. We shall do this from now on.

But you may have wondered why we did not adopt these conventions before, in Section 3-1, or, since we failed to do so at that time, whether conclusions obtained there are not now erroneous.

The answer is that our convention of + and - signs is arbitrary and could be turned around completely, as long as we somehow distinguish in our numerical work between "up" and "down" components. In Section 3-1, where  $v_{0y}$  was kept jealously at zero value, only components of one variety, namely "downward," appeared in any problem; hence there was no need to introduce a differentiating sign convention then.

With this treatment of one type of vector components as a background, we need perhaps only point out that the same sign convention, if applied to *displacement* components, will allow us to solve in a similar manner all problems involving the height of a trajectory, the range, etc. After all, displacements also are vectors in the sense that they can be represented by arrows and can be added (composed) or resolved according to the same simple parallelogram scheme which we used for velocities. For example, our use of the Pythagorean theorem in Eq. (3-3) shows this vector aspect of displacements clearly. Another example might be the graphical derivation of  $s_y$  at, say, point B in Fig. 3-6, as shown in Fig. 3-9.

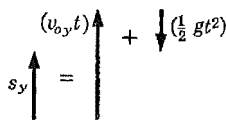


FIG. 3-9.

Therefore  $s_y$  will also attain positive values if represented by an arrow pointing upward, i.e., when the displacement is to a point *above the level of release* of the projectile; and  $s_y$  will assume negative values for displacements measured in the other direction. Then we can use in their original form the pertinent equations involving displacement, such as

$$s_y = v_{0y}t + \frac{1}{2}gt^2, \quad (3-9)$$

provided again that we substitute in all our calculations for  $g$  the value  $-980 \text{ cm/sec}^2$  or  $-32.2 \text{ ft/sec}^2$ .

*Example 4.* The projectile mentioned in the previous examples falls into a valley 1000 ft below the level of the gun. How long was it in the air, and what was the range? Translating, we note that the old data are  $v_0 = 3000 \text{ ft/sec}$  at  $60^\circ$ ,  $\therefore v_{0y} = 2600 \text{ ft/sec}$ ,  $v_{0x} = 1500 \text{ ft/sec}$ ;  $g = -32 \text{ ft/sec}^2$ . To this we add  $s_y = -1000 \text{ ft}$  (a negative sign to indicate that the vertical displacement is to a point *below* the level of release),  $t = ?$   $s_x = ?$

Equation (3-9) can be written:  $\frac{1}{2}gt^2 + v_{0y}t - s_y = 0$ , a quadratic equation easily solved for  $t$ :

$$t = \frac{-(v_{0y}) \pm \sqrt{(v_{0y})^2 - 4(\frac{1}{2}g)(-s_y)}}{2(\frac{1}{2}g)}$$



So far, we have not called in our sign convention. Now we may substitute our values with their proper signs to obtain a numerical answer. (Do this.)\* Having calculated  $t$ , we can solve for  $s_x$ , since it is still true that  $s_x = v_{0x}t$ .

To summarize, we now feel justified in using the following equations of motion in projectile problems, where all symbols have their discussed meaning and the new convention is used to substitute numerical values with correct signs.

$$s_y = \left( \frac{v_{0y} + v_y}{2} \right) t \quad (I) \qquad v_{0x} = v_x = v_0 \cos \theta \quad (VI)$$

$$s_y = v_{0y}t + \frac{1}{2}gt^2 \quad (II) \qquad v_{0y} = v_0 \sin \theta \quad (VII)$$

$$v_y = v_{0y} + gt \quad (III) \qquad v = \sqrt{v_x^2 + v_y^2}, \tan \gamma = \frac{v_y}{v_x} \quad (VIII);$$

$$v_y^2 = v_{0y}^2 + 2gs_y \quad (IV)$$

$$s_x = v_{0x}t = v_x t \quad (V) \qquad s = \sqrt{s_x^2 + s_y^2}, \tan \beta = \frac{s_y}{s_x} \quad (IX)$$

*Conventions regarding sign:*

Vector components in the  $+y$  direction are given  $+$  values.

Vector components in the  $-y$  direction are given  $-$  values.

Consequently,  $s_y$  has  $+$  values if the displacement is above the level of release;  $s_y$  has  $-$  values if the displacement is below the level of release; and  $v_{0y}$  and  $v_y$  have  $+$  values while the corresponding motion is upward, whereas  $v_{0y}$  and  $v_y$  have  $-$  values while the corresponding motion is downward.

The value of  $g$  is at all times taken to be  $-980$  cm/sec<sup>2</sup> or  $-32.2$  ft/sec<sup>2</sup>.

This set is not quite as formidable as it might seem at first glance. We recall that Eq. (I) is really just a definition of average velocity for uniformly accelerated motion, i.e.,  $\bar{v} = (v_{0y} + v_y)/2 = s/t$ . Equation (II) is, on the other hand, an experimental law (from free fall experiments), and therefore, so to speak, on a higher plane. Equations (III) and (IV) can be derived from (I) and (II). Equation (V) is a second experimental law (i.e., the horizontal component of motion  $v_x$  is unaccelerated). The remaining Eqs. (VI)-(IX) simply refer to the mathematical properties of vectors: Eq. (VI) includes the rule

\* At such times it is well to recall that of two possible solutions of such quadratics one may be physically irrelevant or even absurd. (The latter applies if  $t$  comes out negative or if the quantity under the square root is negative.) What does this reveal about the role of mathematics in physical science?

defining how to find the  $x$  component of vector  $v_0$ , while Eq. (VII) provides for the  $y$  component. Equations (VIII) and (IX) give the rule for vector addition (magnitude and direction of resultant, first for the total actual velocity, second for the total actual displacement). Finally, the sign conventions, which quickly become second nature in use, reflect only a commonsensical way of differentiating between addition and subtraction of quantities.

When a practicing physicist has to solve projectile problems, he does not really need to have before him all these defining equations and the conventions. He understands these tacitly; what he will regard as important and essential about ideal projectile motion are Eqs. (II) and (IV), modified as follows:

$$s_y = (v_0 \sin \theta)t + \frac{1}{2}gt^2; \quad s_x = (v_0 \cos \theta)t.$$

He will ordinarily combine these equations by eliminating  $t$  between them (i.e., in the left equation, replacing  $t$  by  $s_x/v_0 \cos \theta$ ), thus obtaining

$$s_y = \tan \theta s_x + \frac{1}{2}g\left(\frac{s_x^2}{v_0^2 \cos^2 \theta}\right). \quad (3-10)$$

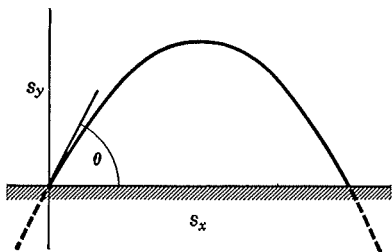


FIG. 3-10. This curve represents Eq. (3-10).

This last equation he will consider to be *the* equation of the trajectory *in vacuo*, and he can attack most problems with this formula (see Fig. 3-10). For us it must suffice to realize that the whole truth about projectile motion, through mathematics, can be so economically expressed and so conveniently handled. Also, we should realize that Eq. (3-10) is of the following form:

$$s_y = (\text{a constant}) \times s_x + (\text{another constant}) \times s_x^2.$$

This once more implies that the *trajectory is parabolic* (no longer simply the lower right *half* of the parabola in Fig. 3-4, which applied where  $\theta = 0$ , whence  $\tan \theta = 0$  and  $s_y = (\text{constant}) \times s_x^2$  only).

**PROBLEM 3-2.** The two curves of Fig. 3-11 differ only in the placement of the origin of the coordinates, so that in (a)  $\theta = 0$ , in (b)  $\theta = 60^\circ$ . Copy the two parabolas on a sheet of graph paper and show by computation for several points that the equation is for (a)  $Y = (\text{a constant}) \cdot X^2$ , for (b)  $Y = (\text{a constant}) \cdot X + (\text{another constant}) \cdot X^2$ . Once more we see that a

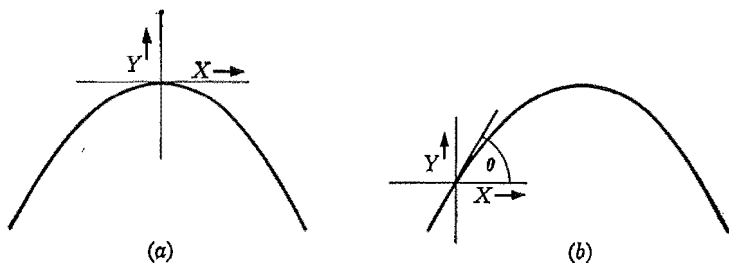


FIG. 3-11.

knowledge of parabolas is all we should really need to solve practical projectile problems. Nevertheless, at our level it will prove easier for us if we continue to solve our problems by means of Eqs. (I)-(IX) rather than, say, by Eq. (3-10) plus the theory of parabolas.

**3-4 Applications of the law of projectile motion.** Let us review what we have done. First, starting from some experimental observations, we have found the separate equations for the horizontal and the vertical components of a general projectile motion. In particular, the horizontal component is found to be unaccelerated, and the vertical one obeys the laws of motion of uniformly accelerated objects. Secondly, the total actual motion (displacement  $s$  and speed  $v$ ) is obtained in direction and magnitude by a vector addition of the components, following an empirically justified procedure of adding by the parallelogram method. Now comes the reward. We see what a wide variety of projectile problems we can solve with this set of equations (which we may regard as the algebraic expression of *the general law of projectile motion*).

*Example 1.* A gun pointed at an angle  $\theta$  above the horizon shoots a projectile with muzzle velocity  $v_0$ . What is the time  $t$  needed for the projectile to return to level ground? That is, given  $v_0$ ,  $\theta$ , and, of course,  $g$ , find  $t$  if  $s_y = 0$  (Fig. 3-12).

*Solution:* Since

$$s_y = v_{0y}t + \frac{1}{2}gt^2, \quad (\text{II})$$

$$0 = (v_0 \sin \theta)t + \frac{1}{2}gt^2,$$

and

$$t = \frac{-2v_0 \sin \theta}{g}.$$

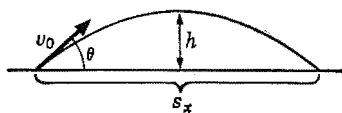


FIG. 3-12.

Since  $g$  takes on a negative value,  $t$  will come out positive, as it surely must.

*Example 2.* In the previous example, what is the maximum height reached? That is, given  $v_0$ ,  $\theta$ ,  $g$ , find  $s_y$  ( $\equiv h$ ) if  $v_y = 0$ .

*Solution:* By Eq. (IV),  $0 = (v_0 \sin \theta)^2 + 2gs_y$ ,

$$\text{and} \quad s_y (\equiv h) = \frac{-(v_0 \sin \theta)^2}{2g}.$$

Again  $s_y$  is positive, since  $g$  has a negative value.

*Example 3.* In the first example, what is the range, i.e., if  $v_0$ ,  $\theta$ , and  $g$  are given, and  $s_y = 0$ , what is  $s_x$ ?

*Solution:*  $s_x = v_{0x}t = (v_0 \cos \theta)t$ . However, since  $t$  is not explicitly known, it must first be calculated, and this can be done as shown in Example 1. (This may be called a two-step problem.) By direct substitution, it follows that

$$s_x = \frac{-v_0^2 2 \cos \theta \sin \theta}{g} = \frac{-v_0^2 \sin (2\theta)}{g}.$$

It should be emphasized again that these results would hold strictly only for projectiles moving in a vacuum above a flat earth. The actual effects of air resistance alone may decrease the effective range by 50%. Evidently, for practical applications, we should have to investigate how such effects modify these results through theoretical and empirical corrections.

**3-5 Galileo's conclusions.** Galileo himself carried his work to the point where he computed fine tables of ranges and heights of trajectories over level ground for different angles  $\theta$ . These calculations proved that for a given initial velocity the range was a maximum if  $\theta = 45^\circ$ , as we can ascertain by inspection of the general equation  $s_x = [-v_0^2 (\sin 2\theta)]/g$ , for  $\sin (2\theta)$  takes on its maximum value, 1, if  $\theta = 45^\circ$ . Galileo also used this result to show that  $s_x$  for a given type of projectile is equally large for any two values of  $\theta$  differing from  $45^\circ$  by an equal amount in each direction (e.g.,  $52^\circ$  and  $38^\circ$ ). Convince yourself of this unexpected fact by substitution, if no simpler way suggests itself. Galileo very penetratingly remarks at this point:

“The force of rigorous demonstrations such as occur only by use of mathematics fills me with wonder and delight. From accounts given by gunners, I was already aware of the fact that in the use of cannon and mortars, the maximum range, that is, the one in which the shot goes farthest, is obtained when the elevation

is  $45^\circ$  . . . ; but to understand *why* this happens far outweighs the mere information obtained by the testimony of others or even by repeated experiment . . . The knowledge of a single effect apprehended through its cause opens the mind to understand and ascertain other facts without need of recourse to experiment, precisely as in the present case, where, having won by demonstration the certainty that the maximum range occurs when the elevation  $[\theta]$  is  $45^\circ$ , the Author [Galileo] demonstrates what has perhaps never been observed in practice, namely, that for elevations which exceed or fall short of  $45^\circ$  by equal amounts, the ranges are equal . . .”

Note the pleasure which he finds in discovering a mathematical law, and that he regards it as the way “to understand *why*” projectiles move as they are observed to do. Note also especially the phrase “without need of recourse to experiment”; Galileo, traditionally regarded as the patron saint of experimental science, clearly warns us that the continual experimental verification of a prediction from a law is unnecessary once the law is sufficiently well established. After the initial mandatory doubts are satisfied, one must sufficiently believe in the law in order to obtain any benefit from it at all.

A second important conclusion from the work on projectiles may seem almost trivial at first glance. In support of the Copernican thesis that the earth moves about its axis and around the sun, Galileo offered an answer to critics who argued that if the earth moved, a stone dropped from a tower would be left behind while falling through the air, and consequently would not land directly at the foot of the tower as observed, but much beyond. Galileo assumes that during the *short* time of fall, the earth, and the top and the foot of the tower may be thought to move forward equally far with uniform velocity. If, then, the whole tower moves with the same speed  $v_{0x}$ , the dropped stone must fall along the tower, because it will retain this “initial” horizontal component of speed, as truly as a parcel dropped from a moving plane lands below the plane or, in Galileo’s analogy, as an object falling from the mast of a moving ship lands at the foot of the mast. From this and equivalent observations concerning the other laws of mechanics has been developed a most valuable generalization, usually called the *Galilean Relativity Principle*: Any mechanical experiment, such as on the fall of bodies, done in a stationary “laboratory” (e.g., on a stationary ship) will come out precisely the same way when repeated in a second “laboratory” (say on a moving ship)

as long as the second place of experimentation moves with *constant velocity* as measured from the first.

To express this more elegantly, let us use the words "coordinate system" instead of "laboratory," since all that counts in the description of experiments is the system of reference used for measurements. Some corner of the rectangular laboratory table might be the origin for all our measurements of distance, and the  $x$ ,  $y$ , and  $z$  directions may correspond to the directions of the three edges. Then we may restate the Galilean Relativity Principle: "All laws of mechanics observed in one coordinate system are equally valid in any other coordinate system moving with a constant velocity relative to the first." If you drop a suitcase in a compartment in a train, be it at rest or speeding at 30 mi/hr or at 60 mi/hr across level terrain, the case will always hit the floor below its point of release, and will fall downwards with the usual acceleration as measured in the car. Consequently, from the motion of a projectile in a car you cannot decide whether the car itself is moving. If the Principle is true in its general form, it follows that no *mechanical* experiment made on our earth can decide such intriguing questions as whether the solar system as a whole proceeds with some steady motion through the space about us.

**3-6 Summary.** Evidently, this has carried us far from the starting point of our discussion. Recall that we started from the simple case of projectile motion in a plane,  $\theta$  being  $0^\circ$  at first. Then we deduced the superposition principles for velocities, learned how to add vector components by the simple parallelogram method, and extended our treatment to the general case, obtaining the *laws of projectile motion* without any restrictions on the initial direction of motion. Finally we heard that even more general principles were induced from these laws. We also noted in passing the power and meaning of the mathematical formulation of physical events, and the multifold predictions flowing from such mathematical laws—whether theoretical, as the determination of a maximum range for  $\theta = 45^\circ$ , or practical, as the construction of gunnery tables.

With these examples of the development and power of physical law in mind, we shall later return to a discussion of "scientific methods." But we shall encounter influences of Galileo's work from time to time throughout the rest of this study, as one might find an ancestral trait running through generations of one family.

For the moment, an urgent task remains: In this section, Part A, we have studied motions of various sorts, but motion apart from the

forces that cause them. To kinematics, we must now add dynamics. These are as two sides of a coin; without both, we cannot purchase the full understanding of any mechanical phenomenon.

### Additional Problems

**PROBLEM 3-3.** What is meant by saying that the paths of projectiles *in vacuo* are parabolic? When we find by experiment that the paths *are* parabolic, what help is that in solving projectile problems? *Why* should trajectories follow parabolic paths instead of perhaps that shown in Fig. 2-1?

**PROBLEM 3-4.** A stone is thrown vertically upwards with an initial speed  $v_{0y} = 100$  ft/sec (Fig. 3-13). Find (at the end of the 1st, 3rd, and 7th second), (a) the displacement  $s_y$  of the stone, (b) the instantaneous velocity. Use your data to plot a  $v$  vs.  $t$  graph (on graph paper) and find from it (*graphically*), (c) when the stone will have a speed of 200 ft/sec (numerically), (d) what the acceleration is at the instant that the stone has reached its topmost position and just begins to turn downwards.



FIG. 3-13

**PROBLEM 3-5.** How long will the stone in Problem 3-4 take to reach again the level of release? Obtain this result first from your graph for Problem 3-4, then check it by independent direct calculation.

**PROBLEM 3-6.** A gun fires a shell at a muzzle velocity of 3000 ft/sec and at an angle of  $30^\circ$  above the horizon. How long will it take to return to earth, how high will the projectile rise, and what is its range on level ground? (Neglect air friction.)

**PROBLEM 3-7.** A baseball is fired by a pitcher horizontally from a vertical height of 6 ft above home plate. It reaches the batter after 0.5 sec. Will the ball be called a strike if the batter does not swing?

**PROBLEM 3-8.** The famous gun known as Big Bertha in World War I had a maximum range on flat territory of about 75 miles. What would have been the muzzle velocity of the projectile if it had exhibited such a range *in vacuo*? (The actual muzzle velocity under ordinary circumstances would be somewhat larger.)

**PROBLEM 3-9.** A baseball is thrown with  $v_0 = 50$  ft/sec at an elevation ( $\theta$ ) of  $60^\circ$ . Show that after 2 sec it has risen 23 ft, has traveled 50 ft horizontally, and moves there at an angle of about  $-39^\circ$  (i.e.,  $39^\circ$  below the horizontal).

**PROBLEM 3-10.** Modern cosmic ray research sometimes employs rockets to sample automatically the radiation at high altitudes. For simple calculation's sake, consider that a self-propelled rocket constantly increases its speed on a straight vertical path so that at an altitude of 18 miles, when the fuel is exhausted, it has attained a speed of 5000 ft/sec. At this point the stabilizers automatically turn the rocket at an angle of  $60^\circ$ , after which the missile continues as an ordinary projectile. What time interval (in

seconds) is available for measurements in the relatively air-free region above 18 miles? What is the total time from the firing of the rocket to the instant it hits the ground? What assumptions have you introduced?

PROBLEM 3-11. A hunter aims his gun barrel directly at a monkey in a distant palm tree. Where will the bullet go? If the animal, startled by the flash, drops out of the branches at the very instant of firing, will it then be hit by the bullet? Explain. Would this last answer hold if the acceleration of gravity were not  $32 \text{ ft/sec}^2$  but only  $\frac{1}{6}$  as great, as on the moon?

PROBLEM 3-12. For each of the following cases sketch two graphs, one of total displacement *vs.* time elapsed, the other of velocity *vs.* time elapsed. Carefully use the convention concerning positive and negative values. (a) A parachutist falls from a plane, after a while opens the 'chute, and floats to the ground. (b) A marble drops from your hand and bounces three times before coming to rest on the floor. (c) A shell is fired at  $60^\circ$  elevation and falls into a deep valley.

PROBLEM 3-13. Read through the Fourth Day (Motion of Projectiles) in the full edition of Galileo's *Dialogues Concerning Two New Sciences*, and on this basis briefly discuss the following: (a) Galileo's examination of the role of air resistance in projectile motion. (b) Galileo's use of experiments to further his argument. (c) Galileo's interest in practical applications of his work.

### Further Reading

Galileo Galilei, *op. cit.* (Ch. 2), Fourth Day. Excerpts in W. F. Magie, *op. cit.*, pp. 19-22; also J. W. Knedler, *op. cit.*, pp. 151-167.



*Part B*

THE STUDY OF FORCES

- 4 Newton's Laws of Motion
- 5 Rotational Motion



*Isaac Newton.*

## THE STUDY OF FORCES

“This subject of the formation of the three laws of motion and of the law of gravitation deserves critical attention. The whole development of thought occupied exactly two generations. It commenced with Galileo and ended with Newton’s *Principia*; and Newton was born in the year that Galileo died. Also the lives of Descartes and Huyghens fall within the period occupied by these great terminal figures. The issue of the combined labours of these four men has some right to be considered as the greatest single intellectual success which mankind has achieved.”

A. N. Whitehead, *Science and the Modern World*.

As we look into history, it would seem that sometimes progress in a field of learning depended on one man’s incisive formulation of the right problem at the right time. It is so with the section of mechanics called *Dynamics*, the science that deals with the effects of forces on moving bodies. The man was Isaac Newton, and the formulation was that of the concepts *force* and *mass*, expounded in three interconnected statements that have come to be called *Newton’s Three Laws of Motion*. These, together with an introduction to rotational motion, form the substance of this Part B, which complements and extends the study of pure kinematics in Part A. In conjunction, these two sections enable us to understand a great many simple mechanical phenomena, and so form an introduction to the next level of problems in physical science.

## CHAPTER 4

### NEWTON'S LAWS OF MOTION

For all the ingenuity and penetration, variety and bulk, of Galileo's work, there is hardly a handful of details in it which today survives in original form as basic elements in contemporary science. Galileo's great contribution was the total effect he had on his contemporaries and immediate successors. But Newton's work is different in this respect. He comes on the stage at the end of the old quarrel whether there should be a new science and what its province might be. The main stream of Newton's brilliant energies therefore is more easily directed to problems *in* science instead of problems *concerning* science; and as a consequence his work, in outline and detail, still is at the basis of introductory treatments of dynamics.

More than that—not only dynamics, but virtually all of our physical sciences depend on the key concept of force, still definable in essentially the same words that Newton first employed in 1687. We shall hear more about his position and person later on, in connection with his discovery of the law of universal gravitation. But now we proceed to a fairly didactic treatment of the famous three Laws of Motion and some of their consequences.

**4-1 Newton's first law of motion.** We may phrase the First Law, or Law of Inertia, as follows:

*Every material body persists in its state of rest or of uniform, unaccelerated motion in a straight line, unless it is compelled to change that state by the application of an external, unbalanced force.*

The essence is this: If you see a moving body deviating from a straight line, or accelerating in any way, then you must assume that a net force (of whatever kind) is acting on the body; here is the criterion for recognizing qualitatively the *presence* of an unbalanced force. But note well that this law does not help you to discover either how large the force is or its origin. There is implied only the definition of force as the "cause" of change of velocity, a definition we already found alluded to in Galileo's work (see Section 2-4). We recall that the Aristotelian scholastics had a rather different view in this matter; they held that force was also the cause of uniform

(unaccelerated) motion. As an example, consider the two opposing opinions that would exist concerning a cart pulled by a horse across a plane in steady motion.

The Galilean-Newtonian commentator would hold that the horse's efforts serve solely to equal and cancel the force of friction on the wheels, so that the *net* force, i.e., the pull of the animal minus the force of friction, is zero, the cart being consequently in a state of equilibrium, a state which by definition includes uniform unaccelerated motion as well as rest. On the other hand, the Aristotelian scholar would explain that since the natural state of the cart is rest, a force has to be supplied by the horse to keep it in uniform motion—even in the absence of friction, if such a circumstance could be approached. We have here not so much a dispute about experimentally observable facts as a difference in conceptualization, complicated by the use of the same word, *force*, in two such disparate meanings. As a matter of fact, the Aristotelian view, here as in some other contexts, is the one closer to contemporary common-sense opinion; since friction is in actuality never absent and is often a very real hindrance in moving an object, it is natural that one develops the intuitive idea that a force is necessary to "keep things moving," and, as a next step, to define force as "the cause of continued motion." The Newtonian turns from this anthropomorphic position to consider the net forces on the moving body, and even dares to cast his definition of force in terms of the *vector sum* (i.e., *net* force) without having first identified the individual components separately. Only the great practical success which followed it could have justified a step that logically is so precarious.\*

Other points are also raised by the first law of motion. It has been made plain that a net force must be supplied to change a body's state from rest to motion or from motion to rest, from one speed to another or from one direction of motion to another even at the same speed. In somewhat unrigorous language still, we may say that material bodies are endowed with a sluggishness, a laziness toward change, an *inertia*. There is very little profit in trying to account for this experimental fact in terms of some physical model or picture—let us simply accept that physical bodies are characterized by inertia just as they are by such other fundamental attributes as volume, chemical constitution, and so forth.

---

\* For a brilliant discussion of the Laws of Motion see *Foundations of Physics* by R. B. Lindsay and H. Margenau. New York: Wiley (1936), pp. 85-92.

If the change of direction of motion implies the action of a net force, we cannot think—as the ancients did—that no force need act on the planets in their orbits to keep their movement circular; on the contrary, we must consider that they are subjected to forces which draw them continually from a straight-line motion.

**PROBLEM 4-1.** Explain in terms of Newton's first law of motion the common experience of lurching forward when a moving train suddenly decelerates and stops. Explain what happens to the passengers of a car that makes a sharp and quick turn. A penny is put on a record, and the motor is started; when the record has reached a certain speed of rotation, the coin flies off. Why?

**PROBLEM 4-2.** Assume that in your laboratory is set up a horizontal, perfectly frictionless table. A block is placed on it and given a small push. What will be the motion of the block? How will this be affected by a straight-line motion of the whole laboratory during the experiment? What if the block moves in an arc sideways? How could you decide whether or not it is your laboratory which is moving in a curve?

**PROBLEM 4-3.** An object weighing  $X$  units hangs on a spring balance (Fig. 4-1) in an elevator. Decide in each of the following cases whether the balance will read more than  $X$  or less, and why: the elevator accelerates upwards; downwards; moves without acceleration down; decelerates on the way down.

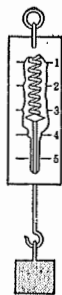


FIG. 4-1.  
Cutaway  
view of a  
spring  
balance.

**4-2 Newton's second law of motion.** So far, we have only a qualitative notion of the concept of force and, incidentally, of the property of bodies called *inertia*. But to develop a science of dynamics we must be able to determine forces and inertias quantitatively, instead of calibrating spring balances arbitrarily to read arbitrary force units, or calling a body's inertia large because, in a qualitative way, a large force is needed to change its state of motion.

This need for quantification is met by Newton's Second Law of motion, which we can phrase as follows:

*The net external (unbalanced) force acting on a material body is directly and linearly proportional to, and in the same direction as, its acceleration.*

Here the point is this: If a net force is qualitatively detected by changes of velocity (First Law), let us define it exactly by the exact rate of change of velocity.\* If we adopt the symbols  $F_{\text{net}}$  and  $a$ , we

\* Of course we remember: any definition is possible—whether the concept as defined is useful must be demonstrated later.

write the Second Law

$$F_{\text{net}} \propto a \text{ (for a given body),}$$

or 
$$\frac{F_{\text{net}}}{a} = a \text{ constant for a given body.} \quad (4-1)$$

The constant so defined is a measure of the body's inertia, for clearly a large ratio of  $F_{\text{net}}$  to  $a$  means that a large force is needed to produce a desired acceleration, which is just what we expect to find true for large, bulky objects, to which we intuitively assign a larger inertia than to small, light objects. If we now symbolize the constant in Eq. (4-1), the measure of inertia, by a letter  $m$  of its own, and give it the alternative name *mass*, we can write

$$\frac{F_{\text{net}}}{a} = m; \quad \text{or} \quad F_{\text{net}} = m \times a. \quad (4-2)$$

Equation (4-2) permits us in principle to assign a numerical value to a net force if we measure the acceleration it produces on a body of known mass, or conversely we could in principle obtain a numerical value for the mass in terms of the acceleration and the net force. But of course you notice the vicious circle: unhappily we find that inertia and force, the two concepts we are trying to quantify rigorously, are mutually interconnected. To find the one we seemingly would have to know the other in advance.

There is one straightforward solution for this apparent dilemma. Before we attempt to measure any masses or forces, we first choose some one unique object, perhaps a certain small piece of polished rock or metal, as the universal *standard of mass*, and regard it arbitrarily as having an inertia of unity, or *unit mass*; for example, name it the *Standard Mass of one gram* (abbreviated 1 gm). Then, by Eq. (4-2), any net force can be evaluated by means of the observable acceleration it produces on our standard object.

This important point will be clearer if we resort to an imaginary, idealized experiment ("thought-experiment"): Place the standard 1-gm object on a smooth horizontal plane where friction is so negligible that we can speak of it as a frictionless plane.\* Now attach to the standard an uncalibrated spring balance (of negligible mass) and pull on it horizontally, as shown in Fig. 4-2. Here the applied force

\* For demonstration purposes, a strikingly close approximation to this ideal is reached when a smooth piece of "dry ice" is placed on a flat, horizontal sheet of plate glass. The thin layer of vapor between the two keeps friction to a minimum.

is indeed the net force. If this net force is constant, the acceleration of the object is also constant, and therefore ascertainable without fundamental difficulties; let us say it comes to  $4 \text{ cm/sec}^2$  for a given pull in this concrete example. But that result tells us at once the magnitude of the pull also, for by Eq. (4-2)

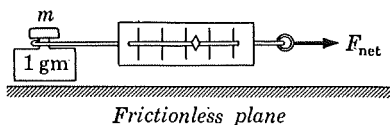


FIG. 4-2. Thought-experiment to define the unit of force.

$$F_{\text{net}} = m \times a, \text{ thus } F_{\text{net}} = 1 \text{ gm} \times 4 \text{ cm/sec}^2 = 4 \text{ gm-cm/sec}^2.$$

The pointer-reading corresponding to that pull may now be marked off\* on the spring balance as  $4 \text{ gm-cm/sec}^2$ ; or, since the unit  $\text{gm-cm/sec}^2$  has traditionally been given the shorter name *dynes*, the marking should be 4 dynes.

A single point will, of course, not serve to fix the calibration of a spring balance. Ideally, therefore, we repeat the whole procedure several times with different applied forces, by observing in each case the corresponding acceleration of the 1-gm object. Eventually we shall thus develop a fully calibrated balance; with its aid we may not only measure other forces directly, but—more important—may reverse the argument to determine quickly the unknown mass  $m_x$  of any other object. For example, the now calibrated balance is applied to  $m_x$  and drawn out to read 15 dynes while accelerating  $m_x$  on the frictionless plane. If the acceleration, measured simultaneously, is  $6.2 \text{ cm/sec}^2$ , then, by Eq. (4-2),

$$m_x = \frac{F_{\text{net}}}{a} = \frac{15 \text{ dynes}}{6.2 \text{ cm/sec}^2} = \frac{15 \text{ gm-cm/sec}^2}{6.2 \text{ cm/sec}^2} = 2.4 \text{ gm}.$$

*In summary*, Newton's second law in conjunction with the essentially arbitrary choice of one standard of mass conveniently fixes the unit of force, permits the calibration of balances, and gives us an operational determination of the mass of all other bodies.

**PROBLEM 4-4.** Recount in detail what steps you must take (in idealized experimentation) to determine the unknown mass  $m_x$  (in grams) of a certain object if you are given nothing but a frictionless horizontal plane, a 1-gm

\* That the force was constant can, strictly speaking, be judged only from the constancy of the observed acceleration. However, happily, there is a separate law of nature governing the behavior of elastic bodies like springs in spring balances, by which the displacement of the spring, and therefore the position of the pointer, is steady and unique for constant forces.

standard, an uncalibrated spring balance, a meter stick, and a stop watch. *Note:* Your answer represents a first approach toward an operational definition of mass, and will prove a great aid in comprehension of dynamical problems.

**PROBLEM 4-5.** What is the mass  $m_x$  if the observed acceleration of the object on the frictionless horizontal plane is  $0.85 \text{ cm/sec}^2$  while the balance reads off a net accelerating force of 22 dynes? What acceleration will a net force of 1.0 dyne produce on a 3.8-gm mass?

**PROBLEM 4-6.** On a certain horizontal plane the friction under a moving body is not zero but 7.4 dynes, while the balance pulls horizontally with 13.6 dynes. If the observed acceleration of the body is  $2.2 \text{ cm/sec}^2$ , what is its mass?

**PROBLEM 4-7.** The mass of an average student is about 75 kgm. As he sits in his car and starts from rest, his acceleration may be about  $200 \text{ cm/sec}^2$ . What force must the seat exert to accelerate him?

**4-3 Standard of mass.** It will be readily appreciated that the Standard of Mass that represents 1 gram, though essentially arbitrary, has been chosen with care. For scientific work, 1 gram was originally defined as the inertia offered by a quantity of  $1 \text{ cm}^3$  of distilled water at  $4^\circ\text{C}$ . This decision, dating from the late 18th century, is rather inconvenient in practice. Although standardizing on a certain amount of water has the important advantage that it permits cheap and easy reproduction of the standard anywhere on earth, yet there are obvious experimental difficulties owing to the effects of evaporation, the additional inertia of the necessary containers, relatively poor accuracy for measuring volumes, and so on. Therefore, it has become accepted custom to use as the standard of mass a certain piece of precious metal, a cylinder of platinum alloy kept under careful guard at the *Bureau Internationale des Poids et Mesures* at Sèvres, a suburb of Paris (alongside the standard of length, a metal bar regarded as representing the distance of 100 centimeters). For use elsewhere, accurate replicas of this International standard of mass have been deposited at the various bureaus of standards throughout the world; and from these, in turn, auxiliary replicas are made for distribution to manufacturers, laboratories, etc.

Again, it was a matter of convenience and accuracy to make this metal block have an inertia 1000 times that of a one-gram mass; and thus the standard object (see Fig. 4-3) is a cylinder  $1\frac{1}{2}$  in. high of mass 1000 gm, called 1 kilogram (abbreviated 1 kgm). Technically, however, the unit which scientists generally work with is still the  $1/1000$ th part of that large kilogram standard, and for all our



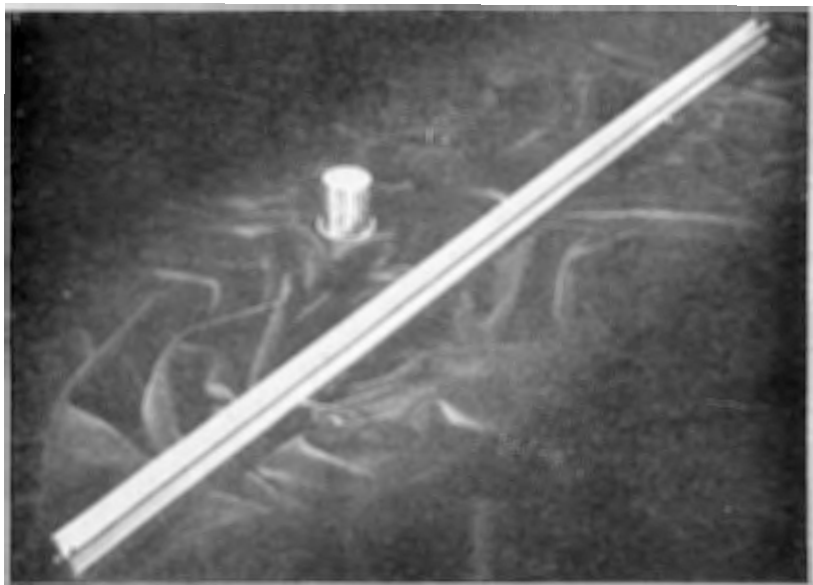


FIG. 4-3. Standard kilogram, together with the standard meter.  
(Courtesy of National Bureau of Standards.)

discussions and thought-experiments we shall speak and think of our standard as being indeed a small piece of metal having the mass of 1 gram. This is about 2.2/1000th of one pound, the unit of mass generally used in the U.S.A. for commercial purposes.\*

**4-4 Weight.** Objects can be acted on by all kinds of forces: by a push from the hand; or by the pull on a string attached to the object; or by a magnetic attraction if the object is made of iron or other susceptible materials; or by the action of electric charges; or by the gravitational attraction that the earth exerts on bodies. But no matter what the origin or cause of the force, its effect is always given by the same equation,  $F_{\text{net}} = ma$ . Newton's second law is so powerful precisely because it is so general, and because we can apply it even though at this stage we may be completely at a loss to understand exactly why and how a particular force (like magnetism or gravity) should act on a body.

\* More strictly, 1 gm = 0.002204622 lb avoirdupois, or 1 lb = 453.5924 gm. But the easily remembered relation 1 lb = 453.6 gm, or even 1 lb = 454 gm, is sufficient for us. For a short and authoritative article on standards and the Bureau of Standards, see L. B. Macurdy's article "Standards of Mass" in *Physics Today*, 4, 7-11 (April 1951).

If the net force is, in fact, of magnetic origin, we might write  $F_{\text{mag}} = ma$ ; if electric,  $F_{\text{el}} = ma$ ; and so forth. Along the same line, we shall use the symbol  $F_{\text{grav}}$  when the particular force involved is the gravitational pull of the earth. Because this case is so frequently considered, a special name for  $F_{\text{grav}}$ , namely *weight*, and a special symbol,  $W$ , is generally used.

Of all the forces a body may experience, the gravitational force is the most remarkable. Let us discuss it in detail. First of all, we are struck by the fact that gravitational forces arise between two bodies (such as a stone in my hand and the earth below) without any need to prepare them for it—simply by placing them near each other. Before two pieces of steel will attract each other by magnetism, we must specially treat at least one of them to magnetize it; and before electric forces can act between two spheres, at least one has to be sprayed with charges. But gravitational pulls are inherently present, although ordinarily the mutual attraction is too small to become noticeable between objects lying on a table. At any rate, we shall concentrate at first on the more spectacular case of the earth's pull on bodies at or near its surface. There,  $F_{\text{grav}}$  acts whether we wish it or no, whereas we can apply or remove at will mechanical, magnetic, and electric influences.

Our second point is this: It is easy to change the mechanical and electric forces on a material particle—by pulling harder on the string to which it is tied, or by depositing more charges on it, etc.—but it is not at all like that with its weight. No matter what we may do to it, a body's weight  $F_{\text{grav}}$  at a given locality is essentially unaffected. To be sure, the farther the two bodies are separated, the less the gravitational attraction between them, as borne out by experiment, so that the weight of an object disappears completely as we remove it to a spot extremely far from all other objects, for example to interstellar space.\* But if *at a certain spot* on this globe the weight of a block of metal is 1000 dynes, nothing will change this result short of destroying the identity of the object by cutting something off or by adding to it. Even if we interpose screens of all kinds between block and earth, the gravitational pull of the earth remains unchanged, as has been proved long ago by careful experimentation.

Presupposed throughout the previous paragraphs was some accurate method for measuring  $F_{\text{grav}}$ . We might simply drop the ob-

---

\* The same result, though for an entirely different reason, is obtained by placing the object—in thought—at the center of our planet, where the gravitational pulls from all sides will fairly well cancel.

ject, allowing  $F_{\text{grav}}$  to pull on it and to accelerate the body in free fall, and then find the magnitude of  $F_{\text{grav}} = \text{mass } (m) \times \text{acceleration owing to gravity } (g)$ . Knowing  $m$  by our previous type of experiment (Section 4-2), and observing experimentally the acceleration  $g$  in free fall, we can at once calculate  $F_{\text{grav}}$ .

Happily there is another method, easier and more direct. We need only our previously calibrated spring balance; from it we hang the body for which  $F_{\text{grav}}$  is to be determined, and then wait until equilibrium is established. Now we do not allow  $F_{\text{grav}}$  to be the only force on the body, but instead balance it out. When the pointer comes to rest—say on the 500-dyne reading—then we know (by Newton's first law) that the upward pull of the spring,  $F_{\text{bal}}$ , just counterbal-

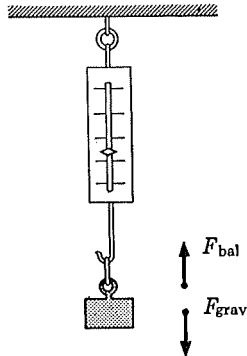


FIG. 4-4. Weighing with a spring balance.

ances the downward pull of  $F_{\text{grav}}$  on the object. The net force is zero. While oppositely directed, these two forces on the same object are numerically equal, and therefore  $F_{\text{grav}}$ , the weight in question, must also be 500 dynes (Fig. 4-4). In passing we note the curious fact that in "weighing" we did not "read off" the weight, but rather the magnitude of an oppositely directed balancing force. This might be called a *static* determination of weight, whereas the previous one was *dynamic*. As a matter of experimental fact, both procedures, though so different, give equal values at the same locality.

Yet another remarkable attribute of the gravitational force is revealed by Galileo's famous observation that (again at a given locality) all objects fall with the same gravitational acceleration  $g$ . This can mean only that the gravitational attraction  $F_{\text{grav}}$  is simply proportional to the mass of the attracted object, that its composition, shape, etc. do not influence  $F_{\text{grav}}$  at all. On reflection, this is an astonishing discovery. The magnitude of other forces (e.g., electric and magnetic ones) is not at all simply proportional to the mass of the affected object. On the contrary; two spheres of equal mass but of different material will generally behave entirely differently when in the presence of the same magnet or electric charge. That there should exist a universal and strictly linear proportionality between

weight and mass is logically as unexpected as if we had found that the weights of objects are proportional exactly to the square root of their volumes or, to use an analogy, if it were discovered that on some far-off planet a man's wealth is proportional to the size of his person. Certainly, wealth and size do not thereby become identical concepts, but such a simple empirical relationship between them should allow us quickly to judge a man's bank account by surveying his bulk, and vice versa. Similarly, although  $F_{\text{grav}}$ , the weight, and  $m$ , the mass, of an object refer to entirely different concepts, the simple relationship between the corresponding values soon puts us in the habit of judging the relative weight from an estimate of the mass, and vice versa.

In summary, the dynamic method of measuring weights by  $F_{\text{grav}} = m \cdot g$  involves a prior determination of mass  $m$  and also a measurement of  $g$ . Now, while  $g$  is constant for all types of objects at a given locality and may for most purposes be taken as 980 cm/sec<sup>2</sup> or 32.2 ft/sec<sup>2</sup> on the surface of the earth,\* the *exact* value is very measurably different at different localities. As Table 4-1 shows, the variation is from about 978.04 cm/sec<sup>2</sup> at the equator to 983.22 cm/sec<sup>2</sup> at the North Pole—no less than  $\frac{1}{2}\%$ . A 1-gm mass will weigh about 981 dynes in Maine and 979 dynes in Florida. Even at the same latitude,  $g$  will decrease with increasing height from sea level; and although the effect is generally quite feeble, so accurate are the methods of measurement that the minute change in  $g$  from one floor of a laboratory to the next is actually discernible!

Table 4-1. Gravitational acceleration

Location	Value of $g$ (cm/sec <sup>2</sup> )
Cambridge, Mass.	980.4
Chicago	980.3
Denver	979.6
Key West	979.0
San Francisco	980.0
Washington, D. C.	980.1
Paris	980.9
Latitude 0°, sea level	978.039
Latitude 90°, sea level	983.217
Sun's surface	27440
Moon's surface	167

\* In some technical work a "standard value" of 980.665 cm/sec<sup>2</sup> is used.

**PROBLEM 4-8.** You are given a frictionless horizontal plane, a 1-gm standard mass, an *uncalibrated* spring balance, a meter stick, stop watches, and assistants. Without further assumptions or aids, and calling only upon Newton's laws of motion, recount in complete detail what steps you should take in idealized experimentation to determine the unknown weight  $F_{\text{grav}}$  (in dynes) of an object in two essentially different ways.

**PROBLEM 4-9.** A standard mass is taken from Paris, France, to our National Bureau of Standards in Washington, D. C. How much is the mass in Washington if it was 1 kgm in Paris? What is its weight in Paris and in Washington? What is the percentage change in its weight? (Consult Table 4-1 for necessary data.)

**PROBLEM 4-10.** At the equator (latitude  $0^\circ$ ), the gravitational acceleration  $g$  is  $978.039 \text{ cm/sec}^2$  at sea level and  $977.630 \text{ cm/sec}^2$  at a height 1000 meters above sea level. If these results were to be confirmed by comparing the weight of a 500-gm object at these two levels on the same spring balance, what must be the magnitude of the smallest division on the scale that one can accurately estimate?

**4-5 The equal-arm balance.** Before we leave the crucial—and initially perhaps troublesome—concept of mass, we must mention a third way of measuring the mass of objects, in everyday practice by far the most favored and accurate method.

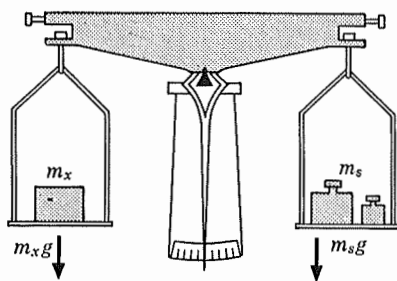


FIG. 4-5. Weighing with an equal-arm balance.

By way of review, recall first that we need the essentially arbitrary standard. Once this has been generally agreed upon, we can calibrate a spring balance by using it to give the standard measurable accelerations on a smooth horizontal plane. Then the calibrated balance can serve to determine other, unknown masses, either (a) by a new observation of pull

and resulting acceleration on the horizontal plane, or (b) by measuring on the balance  $F_{\text{grav}}$  for the object in question and dividing this value by the measured gravitational acceleration  $g$  at that locality.

To methods (a) and (b) we now add method (c), known of course as "weighing" on an equal-arm balance—at first glance seemingly simple and straightforward, but in fact conceptually most deceptive. We place the unknown mass  $m_x$  on one pan (Fig. 4-5), and add a sufficient quantity of marked auxiliary standard masses on the other side to keep the beam horizontal. When this point is reached, the

gravitational pull on the unknown mass, namely  $m_x g$ , is counterbalanced by the weight of the standards,  $m_s g$ . (Note that we assume here the sufficiently verified experimental result concerning the equality of gravitational accelerations for all masses at one locality.) But if  $m_x g = m_s g$ , then  $m_x = m_s$ . Counting up the value of all standard masses in one pan tells us directly the value of the mass on the other side. The whole procedure is much to be preferred over methods (a) and (b), for the range, the accuracy of calibration, and the consistency of a spring balance are all quite limited, whereas in method (c) the equal-arm balance can handle accurately a very large range of masses, and the accuracy of individual measurements (which can be as good as  $10^{-7}$  gm with special care) even without refinements is easily better than 1/1000th of a gram. And the results are repeatable and consistent, there being only these three sound and verifiable assumptions implied: that the arms of the balance are straight and equally long, that the value of  $g$  is the same under each pan, and that the initial marking of the auxiliary standard masses is permanently correct.

We probably take for granted that a result obtained by method (a) is always equal to the result, for the same object, involving methods (b) or (c). And yet, a moment's thought should convince us that this identity is completely unforeseeable—either an incredible coincidence or a symptom of a profound new law of nature. To begin with, consider that the mass was obtained in case (a) in a dynamic measurement: the object actually is allowed to “give in” to the force, is actually accelerated by the pull of the horizontally moving spring balance, and the mass so determined is truly a measure of the inertia of the object. Gravitational forces do not enter at all into the picture; in fact, the experimental measurement could be carried out just as well or better in gravitation-free space. On the other hand, methods (b) and (c), the static procedures of weighing on a vertically held spring balance or on an equal-arm balance, could not be invoked at all in the absence of gravity. Here the whole measurement depends on one property of material bodies: that they are attracted to other objects such as the earth; and inertia plays no role whatever. We see that case (a) on one side and (b) and (c) on the other measure two entirely different attributes of matter, to which we may assign the terms *inertial mass* and *gravitational mass* respectively.

For practical purposes we shall make little distinction between the two types of mass. But in order to remind ourselves how essen-

tial and rewarding it may be to keep a clear notion of the operational significance of scientific concepts, and that, historically, considerable consequences may follow from a reconsideration of long-established facts in a new light, there are these words from Albert Einstein's and Leopold Infeld's book *The Evolution of Physics*.\*

"Is this identity of the two kinds of mass purely accidental, or does it have a deeper significance? The answer, from the point of view of classical physics, is: the identity of the two masses is accidental and no deeper significance should be attached to it. The answer of modern physics is just the opposite: the identity of the two masses is fundamental and forms a new and essential clew leading to a more profound understanding. This was, in fact, one of the most important clews from which the so-called general theory of relativity was developed.

"A mystery story seems inferior if it explains strange events as accidents. It is certainly more satisfying to have the story follow a rational pattern. In exactly the same way a theory which offers an explanation for the identity of gravitational and inertial mass is superior to the one which interprets their identity as accidental, provided, of course, that the two theories are equally consistent with observed facts."

PROBLEM 4-11. It is conceivable that, in another universe, some bodies have only gravitational mass and others only inertial mass. Describe how some everyday objects would behave if made only of one or only of the other type of material.

PROBLEM 4-12. Assume that some gigantic equal-arm balance is firmly set up halfway between the earth and the moon, so that one pan hangs directly above the surface of the earth, the other directly above the *moon's surface*. If an earth-man with a mass of about 80 kgm (about 170 lb) steps on one pan, what must be the mass (in kgm) of a moon-man on the other side to bring the balance to equilibrium?

**4-6 Systems of units.** The system of units we have been using refers to the centimeter, the gram, and the second as the three fundamental standards of measurement. Therefore it is called the *cgs* system (more properly the *cgs absolute system*). But the laws of nature do not, of course, depend on the magnitude of the arbitrarily chosen standards. In particular, we might instead have referred

\* A. Einstein and L. Infeld, *The Evolution of Physics*, p. 36. New York: Simon and Schuster (1942).

throughout to the units used for common daily transactions in the U.S.A. and Great Britain, i.e., the *foot* instead of the centimeter, the *pound avoirdupois* (lb) instead of the gram. The latter corresponds to the mass of 27.692 cubic inches of water at 4°C and ordinary pressure—a quantity of matter having an inertia about 453.6 times that of a 1-gm mass. As in the case of the gram, the pound is expressed by the equivalent inertia of a cylinder of precious metal, this one being kept by the bureau of standards for the United Kingdom in London.\* The relation  $1 \text{ lb} = 453.6 \text{ gm}$  or  $1 \text{ kgm} \doteq 2.2 \text{ lb}$  is a choice that had no particular significance, and is historically rather accidental. At any rate, together with the equivalence between 1 ft and 30.48 cm, it comprises the relation between the cgs system of units and what is called the British Absolute System, or foot-pound-second (fps) system. In a more general sense, the one is called a *metric* system, the other a British system.

Certainly, spring balances might be calibrated by accelerating a 1-lb standard mass, so that  $F_{\text{net}} = m \text{ (in lb)} \times a \text{ (in ft/sec}^2\text{)}$ . The unit of force so obtained, and used in connection with the fps system, is the lb-ft/sec<sup>2</sup>, called the *poundal* (abbreviated pdl), quite analogous to the name *dyne* given to 1 gm-cm/sec<sup>2</sup>, the unit of force used in the cgs (absolute) system. As you should confirm by direct calculation,  $1 \text{ pdl} = 1.383 \times 10^4 \text{ dynes}$ , or  $1 \text{ dyne} = 7.23 \times 10^{-5} \text{ pdl}$ .

A third system of measurement, one which at present is gaining favor among many scientists and may in time replace the cgs system, runs parallel to the latter but uses the kilogram and meter themselves, not fractions thereof, as units of mass and length, hence its name, meter-kilogram-second, or *mks* system of units. The logical unit of force is that net force which accelerates a 1-kgm mass by 1 m/sec<sup>2</sup>, i.e., the kgm-m/sec<sup>2</sup>, for which the name *newton* exists. Clearly,  $1 \text{ newton} = 10^5 \text{ dynes}$ .

Finally, although this recital has actually barely hinted at the length and richness of detail of the full list of different customs in measurements throughout the world, there is a fourth system to be mentioned, the British Engineering System, used widely among engineers in English-speaking countries. Whereas the previous three systems were parallel except for the choice of magnitude of unit mass

\* In the U.S.A., the relation between the pound and the gram (like that between the foot and the meter) is fixed not by a comparison between two metal pieces at the Bureau of Standards at Washington, but by Act of Congress. However, there is no conflict here, for the legal decision was designed to agree with the results of direct comparison of standard masses within the limits of ordinary measurement.



and unit length, this one contains a real conceptual difference. Instead of defining a unit of mass as one of the three fundamental quantities in mechanics, and deriving from it the unit of force, here the mass was originally derived from a unit of force defined *a priori*; specifically, that force is taken as standard which corresponds to the earth's gravitational pull on a one-pound mass, and that force is called a one-pound-force (lbf, also often pound-weight, lbwt, or simply but confusingly, just "pound"). Consequently  $1 \text{ lbf} = 32.2$  poundals. The unit of mass, by the defining equation  $m = F_{\text{net}}/a$ , is the lbf-sec<sup>2</sup>/ft, for which the name *slug* is sometimes used. Table 4-2 gives conversion factors.

Table 4-2. Conversion of Units

SYSTEM	cgs (absolute)	British Absolute	British Engineering	mks
Force	$1 \text{ dyne}$	$= 7.23 \times 10^{-5} \text{ pdl}$	$= 2.247 \times 10^{-6} \text{ lbf}$	$= 10^{-5} \text{ newton}$
	$1.383 \times 10^4 \text{ dynes}$	$= 1 \text{ pdl}$	$= 0.0311 \text{ lbf}$	$= 0.1383 \text{ newton}$
	$4.45 \times 10^5 \text{ dynes}$	$= 32.2 \text{ pdl}$	$= 1 \text{ lbf}$	$= 4.45 \text{ newtons}$
	$10^6 \text{ dynes}$	$= 7.23 \text{ pdl}$	$= 0.2247 \text{ lbf}$	$= 1 \text{ newton}$
Mass	$1 \text{ gm}$	$= 2.205 \times 10^{-3} \text{ lb}$	$= 6.85 \times 10^{-6} \text{ slug}$	$= 10^{-3} \text{ kgm}$
	$453.6 \text{ gm}$	$= 1 \text{ lb}$	$= 0.0311 \text{ slug}$	$= 0.4536 \text{ kgm}$
	$1.459 \times 10^4 \text{ gm}$	$= 32.2 \text{ lb}$	$= 1 \text{ slug}$	$= 14.59 \text{ kgm}$
	$10^3 \text{ gm}$	$= 2.205 \text{ lb}$	$= 0.0685 \text{ slug}$	$= 1 \text{ kgm}$
Accel.	$1 \text{ cm/sec}^2$	$= 0.03281 \text{ ft/sec}^2$	$= 0.03281 \text{ ft/sec}^2$	$= 0.01 \text{ m/sec}^2$
	$30.48 \text{ cm/sec}^2$	$= 1 \text{ ft/sec}^2$	$= 1 \text{ ft/sec}^2$	$= 0.3048 \text{ m/sec}^2$
	$100 \text{ cm/sec}^2$	$= 3.281 \text{ ft/sec}^2$	$= 3.281 \text{ ft/sec}^2$	$= 1 \text{ m/sec}^2$

Hidden behind these figures lies a story with many facets. The multiplicity of existing systems of units points to the arbitrary choice of fundamental quantities, to the process of growth from diverse individual fields, to the role of measurement, and to the operational nature of fundamental concepts. Here, too, we find clues concerning the international character of science, the cooperation and disputes among groups of scientists, the relation between science and industry and between science and government. Even the passion with which certain systems of units are often advocated or attacked helps to reveal a little of the essentially human side of science.

But for the moment we cannot follow these tempting side-roads. As for practical applications, it will save us a great deal of confusion at this stage if we do all our work in the one system in widest use among scientists today, namely the first, the cgs absolute system, in which forces are expressed in dynes, masses in grams, acceleration in

cm/sec<sup>2</sup>. The table for conversion from one system to another will be helpful for expressing all problems and results in a consistent manner. However, those problems in mechanics that involve only *one* of the three concepts—e.g., only forces or only accelerations—are certainly equally well done in any one of the four systems.

**PROBLEM 4-13.** Consider the advantages and disadvantages of each of the four systems. Include considerations of practicality and logical consistency. (Please note, however, that the actual arguments pro and con involve the relative expediency of the systems for use in problems of electricity beyond the level of this account.)

**PROBLEM 4-14.** After ascertaining your exact weight on a scale (in lbf), find your mass in slugs, pounds, kilograms, and grams. Express your weight also in poundals, newtons, and dynes.

**PROBLEM 4-15.** Simply to check on your own ability to handle conversions quickly, find what net force (in dynes) is needed to change the velocity of a 3-lb body from 0.5 meter per second to 50 miles per hour in 2 minutes.

**4-7 Examples and applications of Newton's second law of motion.** Newton's second law, on which were built the previous paragraphs, is so central, so indispensable in physical science, that it will be well worth our looking at a few examples of its application to simple mechanical problems.

*Example 1.* A 4-gm mass  $m$  is pulled across a rough horizontal plane by a force  $F_1$  of 1000 dynes, applied as shown in Fig. 4-6. The frictional force  $f$ , acting as all frictional forces do in a direction contrary to the motion (or tendency of motion) and along the surface of contact, is 800 dynes. What is the acceleration of  $m$ ?

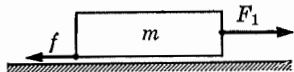


FIG. 4-6

*Solution:* As in all such problems, we make a sketch of the situation and enter on it the significant forces. Since  $F_{\text{net}} = \text{total mass accelerated} \times \text{acceleration}$ ,  $a = F_{\text{net}}/m$ ; but  $F_{\text{net}}$  is here  $(F_1 - f)$ , therefore  $a = (F_1 - f)/m = (1000 - 800) \text{ dynes}/4 \text{ gm} = 50 \text{ cm/sec}^2$ .

Note again that  $(\text{dyne}/\text{gm}) = (\text{cm}/\text{sec}^2)$ , since  $1 \text{ dyne} \equiv 1 \text{ gm cm}/\text{sec}^2$ . If other units had been used, the problem would be solved in the same manner. For example, let  $m = 5 \text{ lb}$ ,  $F_1 = 40 \text{ pdl}$ ,  $f = 10 \text{ pdl}$ . Then  $a = (40 - 10) \text{ pdl}/5 \text{ lb} = 6 \text{ ft}/\text{sec}^2$ , because  $1 \text{ pdl} \equiv 1 \text{ lb ft}/\text{sec}^2$ . Or again, if  $m = 2 \text{ slugs}$ ,  $F_1 = 11 \text{ lbf}$ ,  $f = 6 \text{ lbf}$ , then  $a = 2.5 \text{ ft}/\text{sec}^2$ .

A few words concerning friction may here be interpolated. Experimentally, some remarkable and simple facts exist. To a fair degree of accuracy the friction between the two surfaces of solids may be said to depend only on three details: (1) the normal force  $N$  which presses the surfaces against each other (in our example,  $N$  is simply the weight  $mg$  of the block); (2) the material of which the two surfaces are made; and (3), the type of movement of one object with respect to the other—e.g., whether they *slide* over each other, the case most usually met, or whether the one *rolls* on the other, or whether not quite enough force has been applied to produce relative motion (starting or *static* friction). All this is summarized in the equation form of the law of friction, namely,

$$\text{force of friction } f = \mu N. \quad (4-3)$$

Values for  $\mu$  (Greek letter *mu*), called the *coefficient of friction*, are given in Table 4-3 for a few materials and for the cases of sliding and rolling friction;  $\mu$  for static friction is in every case a little larger than the corresponding value of  $\mu$  for sliding friction.

**Table 4-3. Coefficients of Friction,  $\mu$**

(Approximate values for dry surfaces)

Materials involved	$\mu$ sliding	$\mu$ rolling
Oak on oak	0.25	
Metals on oak	0.4-0.6	
Steel on steel	0.2	0.003
Rubber on concrete	0.7	0.03
Greased surfaces	0.05-0.08	

The normal force  $N$  between the sliding surfaces is by no means always simply the weight  $mg$  of the sliding object. For example, the force  $F_1$  in Fig. 4-7 may not be applied along the horizontal plane but at some angle  $\theta$ ; then the normal force  $N$  between the surfaces is less than  $mg$  by  $F_1 \sin \theta$ , namely, by the upward component of  $F_1$ , that part of  $F_1$  which does not pull the block horizontally but instead tends to lift it vertically. Or again, if  $F_1$  is applied parallel to the surface but the surface itself is inclined at an angle  $\gamma$  (as on

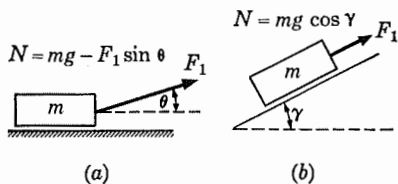


FIG. 4-7.

an inclined plane), the normal force  $N$  between the surfaces is only that component of the weight  $mg$  which acts along a line perpendicular to the two surfaces of contact, namely  $mg \cos \gamma$ . For the present, however, we can confine our attention to those simpler cases where  $\theta$  and  $\gamma$  are zero, i.e., where motion with friction takes place on horizontal planes under horizontally applied forces.

*Example 2.* A block of oak of mass  $m_1 = 20$  gm is dragged across an oak table by a mass  $m_2 = 40$  gm, connected as shown in Fig. 4-8. What is the acceleration of  $m_1$  and  $m_2$ ?

*Solution:* Clearly, the magnitude of the acceleration is the same for  $m_1$  and  $m_2$  if the connecting string does not stretch. The total mass being accelerated is  $(m_1 + m_2)$ ; and the net force acting on this system of two connected masses is the difference between  $F_{\text{grav}}$  on  $m_2$  and  $f$  under  $m_1$ , namely,  $(m_2g - \mu m_1g)$ . Therefore

$$a = \frac{m_2g - \mu m_1g}{m_1 + m_2} = \frac{(40 \times 980 - 0.25 \times 20 \times 980) \text{ dynes}}{(20 + 40) \text{ gm}} = 570 \frac{\text{cm.}}{\text{sec}^2}$$

*Example 3.* A string, thrown over a very light and frictionless pulley, has attached at its ends the two known masses  $m_1$  and  $m_2$ , as shown in Fig. 4-9. Find the magnitude of the acceleration of the masses.

*Solution:* In this arrangement, called Atwood's Machine after the 18th century British physicist who originated it, the net external force on the system of bodies is  $m_2g - m_1g$  (again assuming  $m_2$  to be larger than  $m_1$ ). The total mass being accelerated is  $m_1 + m_2$  and, consequently,

$$a = \frac{(m_2 - m_1)g}{(m_1 + m_2)}, \quad (4-4)$$

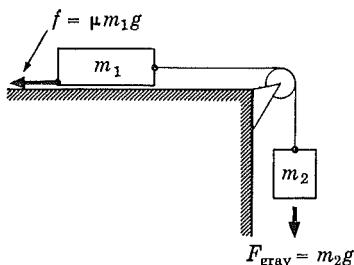


FIG. 4-8.

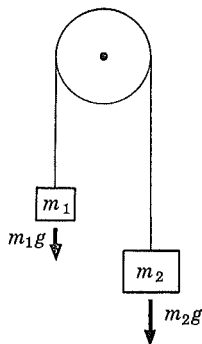


FIG. 4-9. Atwood's Machine.

which we may solve with the known values of  $m_1$ ,  $m_2$ , and  $g$  for that locality. Conversely, if  $m_1$ ,  $m_2$ , and  $a$  are experimentally determined with accuracy, and if corrections are made for the effects of string, pulley, and air resistance, a good value for  $g$  is obtainable.

*Example 4.* An object of mass  $m$  gm hangs from a calibrated spring balance in an elevator (Fig. 4-10). The whole assembly moves upward with a known acceleration  $a$  cm/sec<sup>2</sup>. What is the reading on the balance?

*Solution:* As long as the elevator is stationary, the upward pull of the balance, and hence its reading  $F_1$ , will be equal in magnitude to the weight  $mg$  of the object. The same is true if the elevator is moving *up* or *down* with constant speed, which are also conditions of equilibrium and of cancellation of all forces acting on  $m$ . However, in order to accelerate upward with  $a$  cm/sec<sup>2</sup>,  $m$  must be acted on by a net force of  $ma$  dynes in that direction. In symbol form,

$$F_{\text{net}} = ma; \text{ but } F_{\text{net}} = F_1 - mg,$$

so  $F_1 = ma + mg = m(a + g)$ .

The reading  $F_1$  will be larger than before. Conversely, on decelerating by  $a$  cm/sec<sup>2</sup> while on the way up or accelerating on the way down, the spring balance will relax a little so that there remains on  $m$  some net force ( $mg - F_1$ ) in the downward direction and, of course, of magnitude  $ma$ . Hence the reading  $F_1 = mg - ma$ . Finally, if the elevator were to fall freely ( $a = g$ ), the net force on all objects within must be simply their weight. That is, the last equation becomes  $F_1 = mg - mg$ , i.e., the spring balance reads zero. The object has only one force acting on it, its own weight  $F_{\text{grav}}$ .

Once more we are impressed with the need to speak clearly: A spring balance evidently does not always simply register the weight of an object hanging from it. This would be true only if the whole system is in equilibrium; in other cases, the balance shows a larger or smaller reading depending on the accelerating forces it must supply to the object.

*Example 5.* In the Atwood's Machine problem above, Example 3, we may wish to inquire what the tension is in the string while the masses accelerate, a problem not unlike that of Example 4. First of all we must define clearly the meaning of "tension in a rope": for our purposes it will simply be the force reading if a spring balance

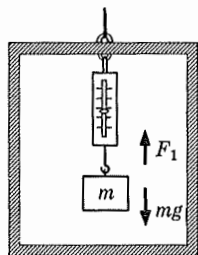


FIG. 4-10. Weight of an object suspended in an elevator.

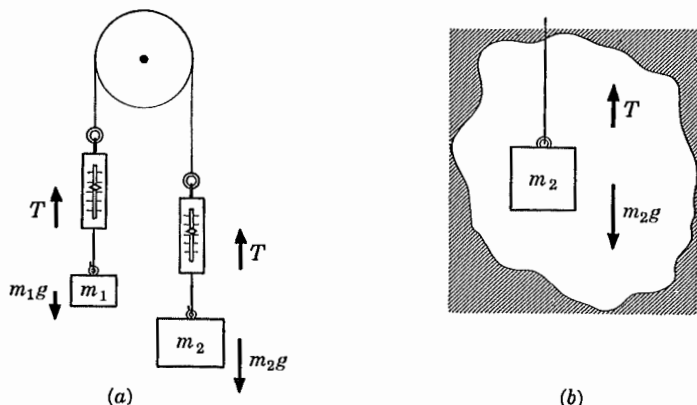


FIG. 4-11.

were cut into the rope [Fig. 4-11(a)]. For a rope of negligible mass (the usual specification in our problem) this force is the same at each end.

To obtain the tension  $T$  in terms of the other data or observables, consider Fig. 4-11(b) closely. If we fasten our attention on one or the other of the two masses only, say  $m_2$ , to the exclusion of the rest of the whole system, then it is evident that  $m_2$  is acted on by a net force ( $m_2g - T$ ); therefore  $(m_2g - T) = m_2a$ , for  $m_2$  is the only mass acted upon by that net force. Rewriting, we obtain

$$T = m_2g - m_2a = m_2(g - a). \quad (4-5)$$

On the other hand, exclusive concentration on  $m_1$  gives  $F_{\text{net}} = T - m_1g$ , or

$$T = m_1a + m_1g = m_1(g + a). \quad (4-6)$$

In answering the question as to the tension in the rope, either of these two results is usable. Incidentally, juxtaposing the two expressions for  $T$ , we obtain  $m_2(g - a) = m_1(g + a)$ , which yields again

$$a = \frac{(m_2 - m_1)}{(m_1 + m_2)}g.$$

This we had already derived on the basis of more direct arguments, which gives us confidence in the method of "isolation" adopted here. Conversely, the last equation may be used to determine the magnitude of the mass on one side of the pulley (say  $m_1$ ) exclusively

by observations made on the other side (i.e., the magnitude of  $m_2$  and  $a$ ). For example,  $m_1 = m_2(g - a)/(g + a)$ .

In a casual way this last example has hinted at a most powerful method of argument in physical science, namely, *the isolation of a specific and small part of the total situation*. As Fig. 4-11(b) symbolized,  $T$  (and from it,  $m_1$ ) could be found by observations made solely on one small portion of the assembly, the rest being perhaps hidden in the inaccessible darkness of a black box. We can thereby tell something about the "inside" of the box though we must remain "outside."

In a sense, nature always confronts the scientist with a "black box problem": he is given some general laws of nature and a few observable manifestations of a complex scheme or hidden mechanism, and his task is to know as much as possible about the scheme or mechanism despite the necessarily scant and incomplete direct information. The problem may be as humble as this one—finding  $m_1$  by observation made on  $m_2$ —or it may be as general as deducing the temperature, motion, size, mass, and composition of a distant star by observations made on a thin beam of its weak light.

**PROBLEM 4-16.** Two objects are joined by one string and dragged across a horizontal oak plank as shown in Fig. 4-12;  $m_1$  refers to a 25-gm mass of oak,  $m_2$  to a 40-gm mass of metal. What must be the pull to the right to give the assembly an acceleration of 5 cm/sec<sup>2</sup>? What is the tension in the rope between  $m_1$  and  $m_2$  in that case?

**PROBLEM 4-17.** On a frictionless horizontal table rests a 15-gm mass  $m_2$  connected as in Fig. 4-13 to hanging masses  $m_1 = 10$  gm and  $m_3 = 25$  gm. Find the acceleration of the system, and the tension in the rope above  $m_1$ .

**PROBLEM 4-18.** A lamp hangs vertically from a cord in a descending elevator. The elevator decelerates by 300 cm/sec<sup>2</sup> before coming to a stop. If during this event the tension in the cord is  $2 \times 10^6$  dynes, what is the mass of the lamp?

**PROBLEM 4-19.** A 160-lb man stands on a platform (spring) balance in a moving elevator. Under what conditions will the balance read  $\frac{1}{4}$  more than when at rest? What is this reading?

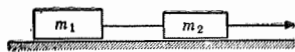


FIG. 4-12.

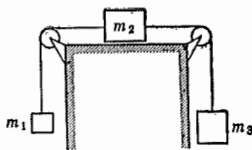


FIG. 4-13.

PROBLEM 4-20. A man tries to obtain a value for  $g$  with an Atwood's Machine but, without knowing it, he and his whole laboratory are freely falling through space. What will be his result for  $g$ , and why?

PROBLEM 4-21. If tomorrow the whole earth were to move with large acceleration in a direction toward the North Star, how would this manifest itself mechanically in different parts of the world?

PROBLEM 4-22. The gun described in Problem 3-6 is in practice found to have a range of only 65% that given by the theory of projectile motion in vacuum. If the shell has a mass of 10 kgm, what is the average horizontal component of the retarding force of friction acting on the shell as it flies through the air? If you make any assumptions in answering this problem, comment on the plausibility of the assumptions.

**4-8 Newton's third law of motion.** Newton's first law defined the force concept qualitatively, and the second law quantified the concept while at the same time providing a meaning for the idea of mass. To these, Newton added another highly original and important Third Law of motion, which completes the general characterization of the concept of force by explaining, in essence, that each existing force has its mirror-image twin. In Newton's words,

"To every action there is always opposed an equal reaction: or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.

"Whatever draws or presses another is as much drawn or pressed by that other. If you press a stone with your finger, the finger is also pressed by the stone. If a horse draws a stone tied to a rope, the horse (if I may so say) will be equally drawn back towards the stone: for the distended rope, by the same endeavour to relax or unbend itself, will draw the horse as much toward the stone, as it does the stone toward the horse, and will obstruct the progress of the one as much as it advances that of the other. . . ."

The point is rather startling: a single particle, all by itself, can neither exert nor experience any force at all. Forces spring up only as the result of interaction of two entities, and then the one pushes or pulls the other just as much as it is being pushed or pulled in return. The earth is attracted upward to a falling apple as much as the falling apple is attracted downward toward the earth. We may choose to call the one force the action and the other the reaction, but the order of naming is really arbitrary. It is not that one of the two forces *causes* the other. They both cause each other simultaneously, as two men, caught in a crowd, may collide and, at the same moment, complain it was the other one who had pushed. In



the same manner, it is as meaningful to say it was the borrowing of some money that caused the debt as it is to reply that the lending of it was the cause of the credit. Action and reaction do stand to each other in the same relation as debit to credit: the one is impossible without the other, they are equally large but in an opposite sense, the causal connection is introduced only artificially, and, most important, they happen respectively to two different bodies.

To emphasize all these points, we might rephrase the third law of motion as follows: Whenever two bodies interact so that body  $A$  applies a force to the other body  $B$  (whether by contact, by gravity, by magnetic interaction, or whatever), then body  $B$  applies simultaneously an equally large and oppositely directed force to body  $A$ .

**4-9 Examples and applications of Newton's third law.** It is easiest to understand the third law in terms of worked-out examples. The first two involve bodies in equilibrium, the last two accelerating systems.

*Example 1.* The simplest case concerns a box (body  $A$ ) standing on the earth (body  $B$ ). Let us enter the forces that act on each. Probably the one that comes to mind first is the weight of the box,  $F_{\text{grav}}$ . We name it here  $F_{1A}$  and enter it as a vertically downward arrow, "anchored" to the box  $A$  at its center of gravity [see Fig. 4-14(a)]. The reaction that must exist simultaneously to this pull of the earth on the box is the pull of the box on the earth, equally large (by the Third Law) and entered as a vertical, upward arrow,  $F_{1B}$ , at the center of the earth in Fig. 4-14(b). This completely fulfills the third law. However, as the second law informs us, if this were the complete scheme of force, the box should fall down while the earth accelerates up. This is indeed what can and does happen while the box drops to the earth, settles in the sand, or compresses the stones beneath it. In short, the

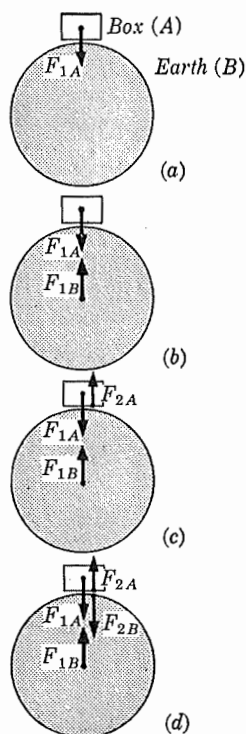


FIG. 4-14. Forces on box and earth in contact.

two bodies do move toward each other until there are built up enough mutual elastic forces to balance the previous set. Specifically, the earth provides an upward push on the box at the surface of contact, as shown in Fig. 4-14(c) by  $F_{2A}$ , an upward arrow "attached" to the box, while, by the law now under discussion, there exists also an equal, oppositely directed force on the ground, indicated by  $F_{2B}$  in part (d) of the figure.

There are now two forces on each body. Equilibrium is achieved by the equality in magnitude of  $F_{1A}$  and  $F_{2A}$  on  $A$ , and by  $F_{1B}$  and  $F_{2B}$  on  $B$ . But beware!  $F_{1A}$  and  $F_{2A}$  are not to be interpreted as action and reaction, nor are  $F_{1B}$  and  $F_{2B}$ . The reaction to  $F_{1A}$  is  $F_{1B}$ ; to  $F_{2A}$ ,  $F_{2B}$ . Furthermore,  $F_1$  and  $F_2$  are by nature entirely different sets of forces—the one gravitational, the other elastic. In short,  $F_{1A}$  and  $F_{1B}$  are equally large by Newton's third law, but  $F_{1A}$  is as large as  $F_{2A}$  (and  $F_{1B}$  as large as  $F_{2B}$ ) by the condition of equilibrium, derived from the *second* law.

*Example 2.* The sketch in Fig. 4-15 involves a system of four elements—a horizontal stretch of earth  $E$  on which a recalcitrant beast of burden  $B$  is being pulled by its master  $M$  by means of a rope  $R$ . Follow these four force couples:  $F_{1E}$  is the push experienced by the earth, communicated to it by the man's heels (essentially by static friction). The reaction to  $F_{1E}$  is the equally large force  $F_{1M}$  exerted on the man by the earth. The man pulls on the rope to the left with a force  $F_{2R}$ , and the reaction to this is the force  $F_{2M}$  with which the rope pulls on the man to the right. A third set is  $F_{3B}$  and  $F_{3R}$ , acting respectively on the donkey and on the rope. Finally, the interaction between earth and animal is  $F_{4E}$  and  $F_{4B}$ . In equilibrium, the separate forces on each of the four objects balance; but if equilibrium does not exist, if the man succeeds in increasing the donkey's speed to the left, then  $F_{3B} - F_{4B} = m_{\text{beast}} \times a$ , and similarly for the other members of the system. And whether there is

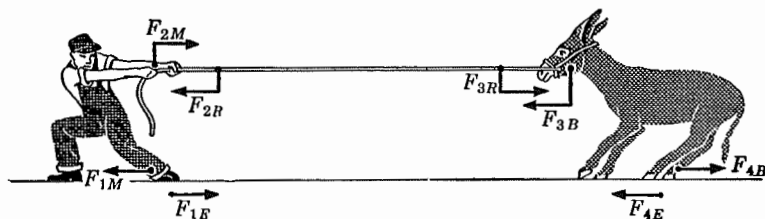


FIG. 4-15. Forces on a system of four objects.

equilibrium or not, any "action" force is equal and opposite to its "reaction."

The whole point may be phrased this way: By the third law, the forces  $F_{1E}$  and  $F_{1M}$  are equal; similarly,  $F_{2M}$  and  $F_{2R}$  are equal. But the third law says nothing whatever about the relationship of  $F_{1M}$  to  $F_{2M}$ , two forces arranged to act on the same body by virtue of the man's decision to pull on a rope, not by any necessity or law of physics. If there happens to be equilibrium, then  $F_{1M}$  will be as large as  $F_{2M}$  by Newton's second law (or, if you prefer, by definition of "equilibrium").

Incidentally, to the master of the beast the difference between zero velocity and a larger, though constant, velocity is of course all that matters; and so he would naturally consider Newtonian physics contrary to common sense if he heard that we lump both these so differently prized states, rest and uniform motion, under one name, equilibrium, and treat them as equivalent. The donkey driver's attitude is certainly understandable, but surely we must avoid it if, as happens again and again, science forces us to break with common-sense preconceptions. The reward is always greater insight into and power over a whole class of natural phenomena.

*Example 3.* Consider an automobile accelerating on a straight road. The force which propels it is essentially the (static) friction between the tires and road at their surfaces of momentary contact, and this force is, by the third law, equal to the contrary force which the earth experiences. Consequently an observer on a fixed outpost in the universe, on a star perhaps, should ideally be in a position to observe the earth moving backward under the car while the latter moves forward over the earth. But evidently these equal forces, if they were the only ones acting, would produce very unequal accelerations. The inertia of the earth being effectively infinitely large compared with that of the car, the former will experience only an infinitely small acceleration. Nevertheless, there is implied an important conceptual clarification: strictly speaking, the  $a$  in  $F_{\text{net}} = ma$ , the acceleration produced by forces on the earth, should be measured not with respect to the earth itself, but rather with respect to some fixed point in space—a "fixed star" would serve. In practice, we do allow ourselves to use the earth as a frame of reference for the measurement of accelerations because the earth's large inertia ensures that it will respond only negligibly to the forces of reaction communicated to it by mutual gravitation, friction, elastic interaction, and so on.

*Example 4.* Consider two blocks at rest on a horizontal, frictionless track (or fitted as cars with light, friction-free wheels). They are connected to each other by a tightly compressed spring (Fig. 4-16). Evidently, the forces  $F_1$  and  $F_2$  on the two blocks owing to their mutual pressure are equal though oppositely directed, and only the balancing forces exerted by a string stretched from one block to the other prevents them from shooting off in opposite directions. Now the string is cut or burned off, and the freed spring propels the cars to the right and the left. At any instant, the opposing forces on the two cars are still equally large; we express this by an equation,

$$F_1 = -F_2.$$

As the spring quickly relaxes more and more, the forces at its two ends will diminish in magnitude until finally the mutual forces on the cars are zero as they separate completely and the spring drops out entirely. Taking an average force  $\bar{F}$  for the whole performance, we have

$$\bar{F}_1 = -\bar{F}_2.$$

However, in the absence of friction,  $\bar{F} = m\bar{a}$  (where  $\bar{a}$  stands for average acceleration), or

$$m_1\bar{a}_1 = -m_2\bar{a}_2. \quad (4-7)$$

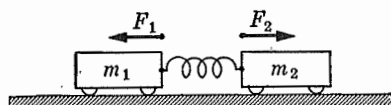


FIG. 4-16. The reaction-car experiment.

The average acceleration is given by  $\bar{a} = (v - v_0)/t$ . For both cars,  $v_0 = 0$  (start from rest), and  $t$ , the time of interaction, is of course the same on each. Thus we rewrite Eq. (4-7):

$$m_1v_1 = -m_2v_2. \quad (4-8)$$

Considering simply the *numerical magnitudes* of the velocities, i.e., neglecting the minus sign (which only serves to remind us that the velocities are toward opposite directions), we note that

$$\frac{m_1}{m_2} = \frac{v_2}{v_1}, \quad \text{or} \quad m_1 = m_2 \frac{v_2}{v_1}. \quad (4-9)$$

In words, these equations reveal that the velocities achieved by two objects by mutually interacting forces are inversely proportional to their masses. Equation (4-9) has special further importance. If one of these objects, say  $m_2$ , is chosen to be a *standard of mass*, then the inertia of any other object  $m_1$  can be directly compared by means of the easily found ratio of measured velocities  $v_2$  and  $v_1$ . Specifically, if in some special case  $m_2 = 1$  gm (by definition), and  $v_2 = 20$  cm/sec and  $v_1 = 5$  cm/sec by observation, then  $m_1 = 4$  gm.

This so-called "reaction-car experiment" immediately offers us a fourth and last method, which we shall call (d) in analogy to the three methods discussed in Section 4-5, for the determination of the mass of an object, namely, by using a standard as mass  $m_2$  and finding the unknown mass from Eq. (4-9). As in method (a), we find here its inertial mass relative to a previously chosen standard; but this new case has the profound advantage that there enters nowhere a calibration of spring balances or even a *measurement* of forces. As long as we know, by ample experimental verification of Newton's third law, that the mutual forces between these two cars are equally large and oppositely directed, there is no further need to investigate their actual magnitudes.

### Additional Problems

PROBLEM 4-23. Make a set of drawings analogous to Fig. 4-14 but involving an apple hanging from a tree.

PROBLEM 4-24. Make a drawing of the situation in Fig. 4-15, but enter the pertinent force pairs for the particular case when man, rope, and donkey *do* accelerate to the left. Represent forces of different magnitudes by arrows of different lengths.

PROBLEM 4-25. Discuss in terms of the Third Law why the tension  $T$  could be assumed to be the same on both ends of the rope in the Atwood's Machines illustrated in Figs. 4-9 and 4-10.

PROBLEM 4-26. A  $3\frac{1}{2}$  ton truck starts from rest and reaches a speed of 40 km per hour in half a minute. What will be the change of speed of the earth during that same time? (Mass of earth =  $6.6 \times 10^{24}$  tons or about  $6 \times 10^{24}$  kgm.) Is the effect on the earth different if the truck takes three times as long to reach the same final speed?

PROBLEM 4-27. The "thought-experiment" involving the reaction cars for the determination of relative masses was first proposed by Ernst Mach, the 19th century Austrian physicist, in one of the early self-conscious examinations of the foundations of mechanics. (His works later stimulated a good deal of modern thought in physics and the philosophy of science.) Describe a detailed step-by-step procedure for using his method to find the inertial mass of an object (in grams). Then describe how you would go on to prove experimentally that the gravitational mass of the object (in grams) has an identical value, and finally, show how you would find the accurate weight of the object (in dynes) at that locality.

### Further Reading

W. F. Magie, *op. cit.* (Ch. 2), pp. 30-46. Excerpts from the *Principia*. Also J. W. Knedler, *op. cit.*, pp. 171-189.

L. T. More, *Isaac Newton, a Biography*. New York: Scribner (1934).

Isaac Newton, *Philosophiæ Naturalis Principia Mathematica*. A. Motte's translation revised by F. Cajori. Berkeley: University of California Press (1934). Read all Prefaces and pp. 1-28, then glance through the rest of this magnificent work.

J. W. N. Sullivan, *Isaac Newton, 1642-1727*. New York: Macmillan (1938).

## CHAPTER 5

### ROTATIONAL MOTION

At this point, we have at our command enough knowledge and tools to deal with a large variety of problems concerning motions and forces. The basic structure has been set up for an understanding of the type of concepts, questions, and methods of answer in the repertoire of the physical scientist. But there is still one dangerous gap to be filled, one main supporting pillar to be moved into place before the next level can be constructed.

In the previous chapters we first acquainted ourselves with the description of uniformly accelerated motions along a straight line, and in particular with that historically so important case of free fall. Next came general projectile motion, an example of motion in a plane considered as the superposition of two simple motions. Then we turned to a consideration of the *forces* needed to accelerate bodies along a straight line. But there exists in nature another type of behavior, not amenable to discussion in the terms that we have used so far, and that is *rotational motion*, the motion of an object in a plane and around a center, acted on by a force that continually changes its direction of action. This topic subsumes the movement of planets, flywheels, and electrons alike.

We shall follow the same pattern as before: concentrating on a simple case of this type, namely, circular motion. We shall first discuss the kinematics of rotation without regard to the forces involved, and finally study the dynamics of rotation and its close ally, vibration.

**5-1 Kinematics of constant circular motion.** Consider a point whirling with constant speed in a circular path about a center  $O$ ; the point may be a spot on a record turntable, or a place on our rotating globe, or, to a good approximation, the planet Venus in its path around the sun.

Before we can investigate this motion, we must be able to describe it. How shall we do so with economy and precision? Some simple new concepts are needed:

(a) The *frequency* of rotation is the number of revolutions made per second (letter symbol  $n$ ), expressed in 1/sec (or  $\text{sec}^{-1}$ ), since the phrase "number of revolutions" does not belong among such funda-

mental physical quantities as mass, length, and time. A conventional type of record, revolving at 78 revolutions per minute (rpm) therefore has a frequency of rotation of  $78/60$  or  $1.3 \text{ sec}^{-1}$ .

(b) Next, we define the concept *period of rotation* (symbol  $T$ ) as the number of seconds needed per complete revolution, exactly the reciprocal of  $n$ , and expressed in units of seconds.

$$T = \frac{1}{n}. \quad (5-1)$$

The record would consequently have a period of rotation of 0.77 sec.

(c) An angular measure is required. The angle  $\theta$  swept through by a point going from  $P_1$  to  $P_2$  (Fig. 5-1) can, of course, be measured in degrees, but it is more convenient at this point to express  $\theta$  by the defining equation

$$\theta = \frac{s}{r}. \quad (5-2)$$

This ratio of arc to radius is a dimensionless quantity; however, the name *radians* is nevertheless attached to this measure of angle, partly to distinguish it from degrees of arc.

To fix the rate of exchange between these two types of angular measure, consider the special case of  $\theta = 360^\circ$ . For that angle,  $s = 2\pi r$ , namely, the circumference of the circle, and therefore  $\theta = 2\pi r/r = 2\pi$  radians. If  $360 \text{ degrees} = 2\pi$  radians,  $1^\circ = 0.0175$  radian and  $1 \text{ radian} = 57.3^\circ$ .

(d) We now inquire into the velocity of a particle executing uniform circular motion. The word "uniform" means, of course, that the rate of spin (the speed  $s/t$ ) does not change. Nevertheless, for future reference it is well to remember that the velocity vector associated with the rotating point does change in *direction* from one instant to the next, although its magnitude, represented by the length of the arrows in Fig. 5-2, is here constant.

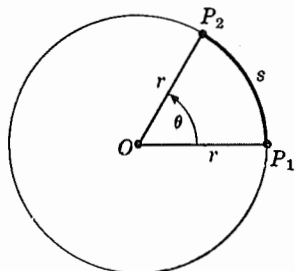


FIG. 5-1. Definition of angle in radian measure:  $\theta = s/r$ .

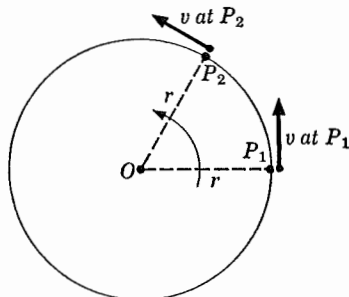


FIG. 5-2.



Let us now concentrate entirely on the magnitude of the velocity, the speed, given again by the ratio of distance covered to time taken; if we know the period or frequency of the motion and the distance  $r$  from the spot to the center of the circle,  $v$  is directly found (usually in cm/sec) by realizing that  $s = 2\pi r$  if  $t = T$ , that is,

$$v = \frac{2\pi r}{T} = 2\pi nr. \quad (5-3)$$

(e) The quantity  $v$  defined in this last equation refers to the magnitude of the tangential or *linear* velocity, that is, to the velocity of the point along the direction of its path. Analogous to this time rate of change of distance stands the powerful concept *angular velocity* [symbolized by the Greek letter  $\omega$  (omega)], which is the time rate of change of angle. By definition, for the type of motion we are here considering,

$$\omega = \frac{\theta}{t}, \quad (5-4)$$

a quantity to which we may assign units of  $\text{sec}^{-1}$  or, more colloquially, radians/sec. If we happen to know  $n$  or  $T$ , the magnitude of the angular velocity is found from the fact that  $\theta = 2\pi$  if  $t = T$ , or

$$\omega = \frac{2\pi}{T} = 2\pi n. \quad (5-5)$$

The formal relation between  $\omega$  and  $v$  is evident by comparing Eqs. (5-3) and (5-5):

$$v = \omega r. \quad (5-6)$$

But direct inspection of the defining equation (5-4) for  $\omega$  reveals a significant difference between  $v$  and  $\omega$ . Every point on a rigid rotating body (e.g., a disk), no matter where it is placed on that body, has the *same*  $\omega$  at the same instant, whereas those different points take on very different values for  $v$ , depending on their respective distances  $r$  from the axis of rotation. The concept of angular velocity  $\omega$  therefore introduces into problems of rotational motion the possibility of constancies not attainable directly in terms of  $v$  alone. There is no better reason for coining a new concept in any field of science.

**PROBLEM 5-1.** How great an angle (in radians and in degrees) is covered by a spot on a 78-rev/min record in 5 minutes of constant rotation? What are the linear speeds of two points, respectively 3 and 12 inches from the center of rotation? What are the corresponding angular speeds?

**PROBLEM 5-2.** Through what total distance does the needle travel while playing a 12" long-playing ( $33\frac{1}{3}$ -rev/min) record? (You must make and explicitly state certain reasonable assumptions in this problem.)

**PROBLEM 5-3.** Find the frequency of revolution, the linear velocity, and the angular velocity for a spot on the equator (earth's radius = 3960 miles; rate of rotation about its own axis = 23 hr 56 min per rotation).

**PROBLEM 5-4.** Find the approximate linear speed of our planet in its yearly path around the sun. (Average distance to the sun =  $93 \times 10^6$  miles.)

**5-2 Vector algebra.** It has been noted that motion with constant speed around a circle implies that the velocity vector is continually changing in direction though not in magnitude. A seemingly paradoxical situation arises at this point and demands a brief mathematical digression: in order to change the direction of a vector quantity at constant magnitude, one must, in general, add to it another vector with far from negligible magnitude.

Consider as an example the case of a disabled ship, adrift in wind and tide at some velocity  $v_1$ , perhaps 2 knots due North. Now a rescue boat throws a line and pulls to give the ship a new velocity  $v_2$ , again 2 knots but now due West. What is the velocity  $v_3$  which the rescue vessel superposes on  $v_1$  to modify it to  $v_2$ ? Figure 5-3 gives the solution in familiar graphic form:  $v_2$  is to be the resultant of  $v_1$  and  $v_3$ , but you notice that the case is a little different from the one in Chapter 3, Fig. 3-8, where we were given  $v_3$  and  $v_1$  (or  $v_a$  and  $v_b$ ) and asked to find  $v_2$  (or  $v$ ). Now the unknown is not the resultant, but one side of the vector parallelogram. Accordingly, we enter first  $v_1$ , then  $v_2$ —both are known in direction and magnitude—and then the connecting line from the head of the arrow representing  $v_1$  to the head of  $v_2$ . Thus the direction and magnitude of the missing side are found, and by measurement from a scale drawing we should find that  $v_3$  in this example is about 2.8 knots to SW.

Comparison of parts (c) and (d) of Fig. 5-3 indicates that the last stage (d) is really superfluous; the connecting line in (c) ade-

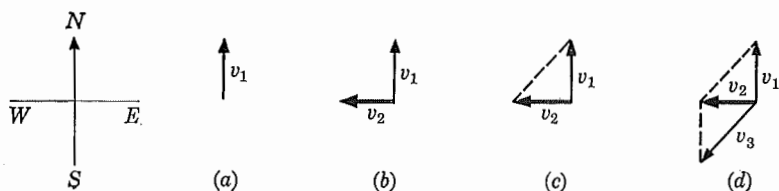


FIG. 5-3.

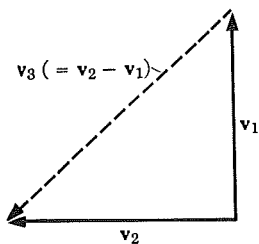


FIG. 5-4. Vector subtraction:  
 $v_3 = v_2 - v_1$ .

quately represented  $v_3$  even before the parallelogram was completed (Fig. 5-4). From now on we may find the difference of any two vector quantities  $A$  and  $B$  by this abbreviated graphical procedure.

Furthermore, we shall occasionally follow the convention of using bold face letters if they refer to vector quantities (i.e.,  $\mathbf{F}$ ,  $\mathbf{v}$ ,  $\mathbf{a}$ ). The addition of two vector quantities  $\mathbf{P}$  and  $\mathbf{Q}$  may then be written  $\mathbf{P} + \mathbf{Q} = \mathbf{R}$ . On reading such equations, we immediately realize that, in general,  $\mathbf{R}$  is not obtained simply by adding the magnitudes of  $\mathbf{P}$  and  $\mathbf{Q}$ , but that one must find  $\mathbf{R}$ , in direction as well as magnitude, by means of a graphical or a trigonometric solution (see Fig. 5-5). In the same manner, the equation  $\mathbf{B} - \mathbf{A} = \mathbf{C}$  refers to a vector subtraction such as is graphically and trigonometrically\* presented in Fig. 5-6.

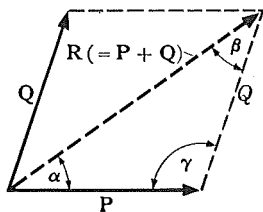


FIG. 5-5. Vector addition, graphically and trigonometrically.

$$\mathbf{R} = \mathbf{P} + \mathbf{Q}.$$

Magnitude of  $\mathbf{R}$ , from Law of Cosines:

$$R = \sqrt{P^2 + Q^2 - 2PQ \cos \alpha}.$$

Direction of  $\mathbf{R}$ , from Law of Sines,  
i.e., from

$$\frac{\sin \alpha}{Q} = \frac{\sin \gamma}{R} = \frac{\sin \beta}{P}.$$

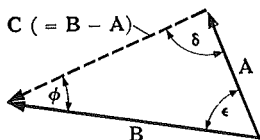


FIG. 5-6. Vector subtraction, graphically and trigonometrically.

$$\mathbf{C} = \mathbf{B} - \mathbf{A}.$$

Magnitude of  $\mathbf{C}$ , from Law of Cosines:

$$C = \sqrt{A^2 + B^2 - 2AB \cos \delta}.$$

Direction of  $\mathbf{C}$ , from Law of Sines,  
i.e., from

$$\frac{\sin \phi}{A} = \frac{\sin \epsilon}{C} = \frac{\sin \delta}{B}.$$

\* See Appendix for a summary of the elements of trigonometry.

*Example of vector addition.* Two men are pulling on ropes attached to a heavy object; the first pulls with force  $F_1 = 9.0 \times 10^6$  dynes ( $\doteq 20$  lbf) at  $40^\circ$  East of North, the second with force  $F_2 = 4.5 \times 10^6$  dynes ( $\doteq 10$  lbf) in a direction  $30^\circ$  West of North. Find the resultant of these two forces, i.e., the vector sum, the single equivalent force  $\mathbf{R}$  which might just as well be applied to this object instead of  $\mathbf{F}_1$  and  $\mathbf{F}_2$ .\*

*Solution:* First, we make a sketch in terms of the chosen arrow symbolism of the relative magnitude and placement of vectors  $\mathbf{F}_1$  and  $\mathbf{F}_2$  [Fig. 5-7(a)]. Then we complete the parallelogram; the diagonal corresponds to  $\mathbf{R}$ , the vectorial sum of  $\mathbf{F}_1$  and  $\mathbf{F}_2$ . The right half of the parallelogram is a triangle composed of  $\mathbf{R}$ ,  $\mathbf{F}_1$ , and a side corresponding to  $F_2$ ; the angle opposite  $\mathbf{R}$ , indicated as  $\gamma$ , is directly given by  $\gamma = 180^\circ - (\theta_1 + \theta_2)$ , in this case  $180^\circ - 70^\circ$ , or  $110^\circ$ .\*\* Now we find the magnitude of  $\mathbf{R}$  by

$$\begin{aligned} R &= \sqrt{F_1^2 + F_2^2 - 2F_1F_2 \cos \gamma} \\ &= \sqrt{20.3 \times 10^{12} + 81 \times 10^{12} - 2 \times 4.5 \times 10^6 \times 9.0 \times 10^6 \times (-0.34)} \\ &= 11.4 \times 10^6 \text{ dynes } (\doteq 26 \text{ lbf}). \end{aligned}$$

The direction of  $\mathbf{R}$  may be fixed in several equivalent ways, but here it would seem most consistent to determine  $\alpha$ , then  $(\theta_2 - \alpha)$ , the angle of  $\mathbf{R}$  with respect to the North direction:

$$\sin \alpha = F_2 \frac{\sin \gamma}{R} = 4.5 \times 10^6 \times \frac{0.94}{11.4 \times 10^6} = 0.37; \alpha \doteq 22^\circ,$$

and

$$(\theta_2 - \alpha) = 18^\circ.$$

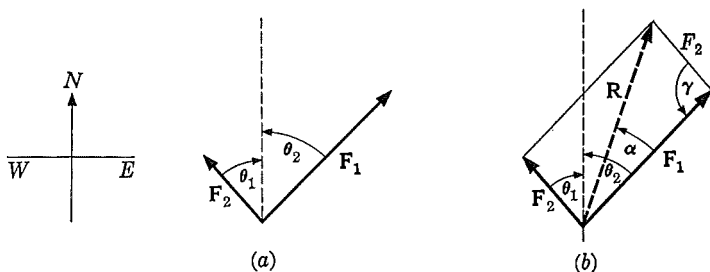


FIG. 5-7. Vector addition.

\* Our previous study of adding velocities to obtain the true velocity of a projectile and of adding forces to find the net force has illustrated the general truth that several superposed vector quantities provide just the same effect as a *single* vector quantity equal to the vectorial sum of the superposed quantities.

\*\* Recall that for angles between  $90^\circ$  and  $180^\circ$  the following relations apply:  $\sin x = \sin (180^\circ - x)$ ;  $\cos x = -\cos (180^\circ - x)$ . Therefore  $\sin 110^\circ = \sin 70^\circ = 0.94$ ;  $\cos 110^\circ = -\cos 70^\circ = -0.34$ .

The answer to the original question, then, is that the resultant force is  $11.4 \times 10^6$  dynes, along a direction  $18^\circ$  East of North. And incidentally, the same procedure is applicable for calculating the sum of *any* two vectors, no matter at what angle to each other, no matter what their physical nature. Even the specific cases where a right angle exists between the two vectors are, in substance, the same problem simplified by the convenient facts that there the factor corresponding to  $2F_1F_2 \cos \gamma$  under the square root is zero (since  $\cos 90^\circ = 0$ ), and that  $\sin \alpha = F_2 \sin \gamma / R = F_2 / R$  (since  $\sin 90^\circ = 1$ ).

Just as examination of Fig. 5-3 leads us to a simplification of the parallelogram method of vector subtraction, namely to Fig. 5-4, so will a moment's thought prove that one may, if one so desires, similarly abbreviate the process of vector addition. Both in Fig. 5-5 and Fig. 5-7(b) only the right half of the parallelogram is really needed; the left half is a congruent triangle, and therefore superfluous for our construction. Therefore, to add two vector quantities, we draw their arrowlike representation and join them head to tail [Fig. 5-8(a),(b)], the resultant  $\mathbf{R}$  then being represented by the arrow running from the tail of the first-drawn to the head of the last-drawn arrow. When there are more than two vectors to be added, we may either do it by adding two of them, then adding a third one to their resultant, then a fourth to *that* resultant, and so forth; or more conveniently, though less accurately, this may be done graphically by adding the corresponding arrows head to tail, as in Fig. 5-8(c). This is generally referred to as the *polygon method* of vector addition; it will soon prove to be an important tool in our argument.

PROBLEM 5-5. Note, describe, and justify carefully the difference between abbreviated vector subtraction (Fig. 5-4) and vector addition [Fig. 5-8(a)].

PROBLEM 5-6. (a) Prove, by graphical construction, that the polygon method of vector addition must give the same result as adding several vectors one at a time.

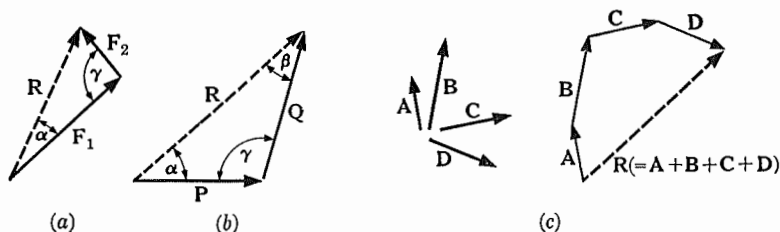


FIG. 5-8. Vector addition.

(b) Prove by graphical construction that in Fig. 5-8(c) the same resultant is obtained regardless of the *order* in which the individual arrows are joined together.

PROBLEM 5-7. (a)  $\mathbf{F}_1 = 5$  lbf due West,  $\mathbf{F}_2 = 8$  lbf at  $30^\circ$  S of E. Find the resultant force, in direction and magnitude, by the parallelogram method (graphically), by trigonometric calculation, and by the abbreviated graphical method (polygon method, which here becomes simply a triangle construction).

(b) To the two forces in part (a), a third is added:  $\mathbf{F}_3 = 4$  lbf  $10^\circ$  W of N. Find the resultant of all three forces by the graphical polygon method (in two different sequences of joining the arrows), and then by adding  $\mathbf{F}_3$  to the resultant of  $\mathbf{F}_1$  and  $\mathbf{F}_2$  in a trigonometric calculation.

*Example of vector subtraction.* A bullet is shot in a direction  $40^\circ$  above the horizontal plane and with an initial velocity  $v_1 = 3.5 \times 10^4$  cm/sec. After 46 sec the bullet returns to the same plane with a velocity  $v_2$  having the same magnitude as  $v_1$ , namely,  $3.5 \times 10^4$  cm/sec, but directed  $40^\circ$  below the horizontal. What was the change of velocity  $\Delta v$  during that time interval (magnitude and direction)?

*Solution:* Again, first make a sketch of the physical situation [Fig. 5-9(a)], and then a sketch of the scheme previously discussed for the subtraction of vector quantities [Fig. 5-9(b)]. Here the angle opposite the side representing  $\Delta v$  is  $(\theta_1 + \theta_2)$ , or  $80^\circ$ ; therefore

$$\begin{aligned}\Delta v &= \sqrt{v_1^2 + v_2^2 - 2v_1v_2 \cos(\theta_1 + \theta_2)} \\ &= \sqrt{12.25 \times 10^8 + 12.25 \times 10^8 - 2 \times 3.5 \times 10^4 \times 3.5 \times 10^4 \times 0.174} \\ &= 4.5 \times 10^4 \text{ cm/sec.}\end{aligned}$$

The direction of  $\Delta \mathbf{v}$  is "obviously" vertically downward, but it is important that this be confirmed rigorously. In Fig. 5-9(c), angle  $(\theta_2 + \phi)$  gives the inclination of  $\Delta v$  with the vertical;  $\theta_2$  is known,  $\phi$  can be easily computed (from  $\sin \phi = v_1 \frac{\sin(\theta_1 + \theta_2)}{\Delta v}$ ), and the sum,  $\theta_2 + \phi$ , is thereby confirmed to be  $90^\circ$ ; i.e.,  $\Delta v$  is perpendicular to the horizontal plane.

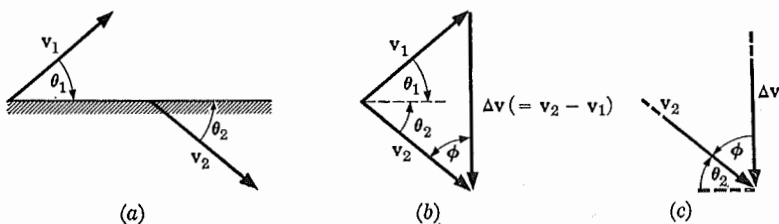


FIG. 5-9.  $\Delta v = v_2 - v_1$ .

This problem, which incidentally was designed to illustrate again the result of a change in direction of a vector even at constant magnitude, can be carried one important step further. Knowing now that during 46 sec the change in velocity was  $4.5 \times 10^4$  cm/sec, we at once can find the average acceleration. By definition,  $a = \Delta v/t$ , therefore  $a = (4.5 \times 10^4/46)$  cm/sec<sup>2</sup> = 980 cm/sec<sup>2</sup>, in fact the expected acceleration of gravity. Furthermore, acceleration being a vector, it is necessary to specify its direction, which is that of  $\Delta \mathbf{v}$ , namely, vertically downward.

**PROBLEM 5-8.** A sailboat cruises on a river which flows in a direction 15° North of East. If there were no wind, the river current would impart to the boat a speed of 5 mi/hr in that same direction. On the other hand, without current, the wind alone would move the boat at 9 mi/hr due South East. Find (a) the resultant of these two vectors, and (b) the difference between these two vectors. The result of (a) represents the total velocity of the boat referred to some fixed point on shore. What is the physical significance of the other result?

**PROBLEM 5-9.** Two forces,  $\mathbf{F}_1 = 10$  pdl,  $\mathbf{F}_2 = 6$  pdl, act on the same object at a 30° angle with each other. What is the resultant net force? What third force must be applied to establish equilibrium? (The latter force is called the *equilibrant*.)

**5-3 Centripetal acceleration.** The results obtained in the previous interlude on vector subtraction are now to be applied to an important problem in rotational motion: What is the change of velocity during uniform circular motion? The answer, to be derived, will be as follows: a body in motion with constant linear speed  $v$  around a circle of radius  $r$  continually experiences an acceleration  $a_c$  toward the center of rotation, where  $a_c = v^2/r = \omega^2 r$ . And being centrally directed at all times,  $a_c$  may be called the center-seeking or *centripetal* acceleration, a term proposed by Newton himself; in all probability he had a knowledge of the essential relationship as early as 1666, although his publication of it, in 1687, is 14 years later than the published derivation of his great contemporary, Christian Huygens. Why the concept of centripetal acceleration is so exceptionally important will appear in the discussion on the law of universal gravitation, where it holds a key position in the arguments that, historically, lead to the universal recognition both of the genius of Newton and of the full power of physical inquiry.

To proceed to a proof of the proposition  $a_c = v^2/r = \omega^2 r$ : In Fig. 5-10, a point moves uniformly through angle  $\theta$  from  $P_1$  to  $P_2$  along the arc  $s$ . The speeds at  $P_1$  and  $P_2$  are equal, but the direction of the velocity vectors has changed through angle  $\theta$ . The change of velocity,  $\Delta v$ , is obtained in the usual manner in Fig. 5-10(b). Now note that the triangle there, and the one contained by  $P_1OP_2$  in Fig. 5-10(a), are similar isosceles triangles. Therefore  $\Delta v/v = x/r$  and  $\Delta v = vx/r$ . On dividing both sides by  $\Delta t$ , the time interval needed for this motion, we obtain  $\Delta v/\Delta t = vx/r\Delta t$ . The left side represents the average acceleration  $\bar{a}$  during  $\Delta t$ , and if, as in Chapter 1, we restrict  $\Delta t$  to smaller and smaller values, in the limit as  $\Delta t$  approaches 0, the average acceleration becomes equal to the instantaneous acceleration  $a$ :

$$a = \lim_{\Delta t \rightarrow 0} \left( \frac{\Delta v}{\Delta t} \right) = \lim_{\Delta t \rightarrow 0} \left( \frac{vx}{r\Delta t} \right).$$

At the same time, however, as  $\Delta t$  (and with it,  $\theta$ ) decreases, the line  $x$  becomes more and more nearly equal to the arc  $s$ , so that in the limit,  $x = s$ , and we write

$$a = \lim_{\Delta t \rightarrow 0} \left( \frac{v}{r} \frac{s}{\Delta t} \right).$$

We note finally that as  $\Delta t \rightarrow 0$ ,  $s/\Delta t \rightarrow v$ , the instantaneous velocity of the moving point, so that

$$a = \frac{v^2}{r}.$$

Earlier we found  $v = \omega r$ ; thus  $a$  is also expressed by

$$a = \omega^2 r.$$

It remains to be shown that the acceleration  $a$  is always center-directed. This is perhaps most directly seen from Fig. 5-10, where it is evident that the direction of  $\Delta v$  will at all times be perpendicular to  $x$  and therefore ultimately along the direction of the radius, toward

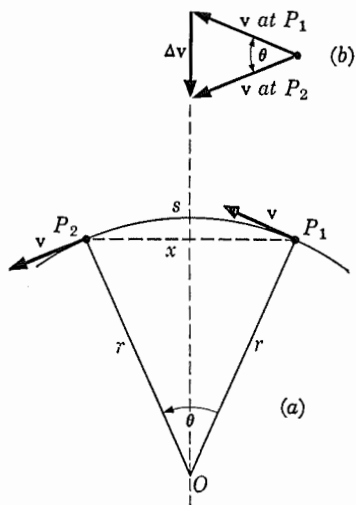


FIG. 5-10.



the center  $O$  of the circle. This justifies placing a small subscript  $c$  after  $a$  to signify a centripetal acceleration, and we have now

$$a_c = \frac{v^2}{r} = \omega^2 r. \quad (5-7)$$

There is often something initially implausible about the concept. The difficulty is to convince oneself that the centripetal acceleration in circular motion is truly at every instant perpendicular to the direction of motion, and therefore does not affect the speed of the object but serves only to change the direction of motion from moment to moment. Without centripetal acceleration, the vector  $\mathbf{v}$  should be constant, i.e., the object would proceed along a straight line.

*Example.* Find the centripetal acceleration of an object on the equator.

*Solution:* This object makes a circular path, covering an angle of  $2\pi$  radians per day ( $= 8.6 \times 10^4$  sec);  $\omega = (2\pi/8.6 \times 10^4)$  rad/sec  $= 7.3 \times 10^{-5}$  rad/sec. The radius of the earth is about  $6.4 \times 10^8$  cm. Therefore  $a_c = \omega^2 r \doteq 3.4$  cm/sec<sup>2</sup>.

**PROBLEM 5-10.** Find the centripetal acceleration owing to the earth's rotation at a place with Latitude  $30^\circ$ . Find the centripetal acceleration for the whole earth, owing to its yearly swing around the sun. (Approximate distance  $= 1.5 \times 10^{13}$  cm; the assumption of a circular orbit is here necessary, though not entirely correct.)

**PROBLEM 5-11.** A good ultracentrifuge may whirl the test sample at a distance of about 0.6 cm from the axis of rotation and at 60,000 rev/min. Find the corresponding centripetal acceleration.

**5-4 Centripetal force.** Accelerations, including centripetal ones, do not occur except through the action of forces. By Newton's second law we must necessarily allow this consequence: A body of mass  $m$  in uniform circular motion is continually acted upon by a net force  $F_c$  of magnitude  $ma_c$ , directed in the same sense as  $a_c$ , i.e., toward the center of rotation; this force, whatever its origin in specific cases, is called a centripetal force. Newton's discussion in the *Principia* of 1687 is very clear:

“DEFINITION: A centripetal force is that by which bodies are drawn or impelled, or in any way tend, towards a point as to a centre. Of this sort is gravity, by which bodies tend to the centre of the earth; magnetism, by which iron tends to the lodestone; and that force, whatever it is, by which the planets are perpetually drawn aside from the rectilinear motions, which otherwise they would

pursue, and made to revolve in curvilinear orbits. A stone, whirled about in a sling, endeavours to recede from the hand that turns it; and by that endeavour, distends the sling, and that with so much the greater force, as it is revolved with the greater velocity, and as soon as ever it is let go, flies away. That force which opposes itself to this endeavour, and by which the sling perpetually draws back the stone towards the hand, and retains it in its orbit, because it is directed to the hand as the centre of the orbit, I call the centripetal force. And the same thing is to be understood of all bodies, revolved in any orbits. They all endeavour to recede from the centres of their orbits; and were it not for the opposition of a . . . force which restrains them to, and detains them in their orbits, which I therefore call centripetal, would fly off in right lines, with an uniform motion . . . Nor could the moon, without some such force, be retained in its orbit. If this force was too small, it would not sufficiently turn the moon out of a rectilinear course: if it was too great, it would turn it too much, and draw down the moon from its orbit towards the earth. It is necessary, that the force be of a just quantity, and it belongs to the mathematicians to find the force, that it may serve exactly to retain a body in a given orbit, with a given velocity . . .”

When a stone—to choose Newton’s concrete example—on a string is made to revolve, it is clear that it is *not in equilibrium*, but that a centrally directed force acts on it, namely the tension in the string, traceable in turn to the pull by the hand. If the string were to break, however,  $F_c$  would cease, and with it also the centripetal acceleration; consequently the direction of the velocity vector at the instant of breaking would no longer be modified and, instead, the stone, now in equilibrium (except for gravity) by Newton’s first law, would fly off in a straight line tangent to the circle.

By the third law, the force  $F_c$  on the stone (while the string is still unbroken) is attended by a reaction, a contrary force with which the stone pulls on the

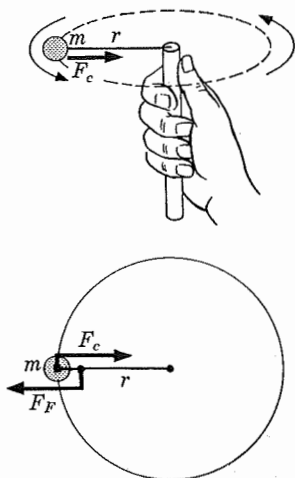


FIG. 5-11. A centripetal force acts on the stone, while a centrifugal force acts on the string and hand.

string and ultimately on the hand at the center. This reaction is named the center-fleeing or *centrifugal* force ( $F_F$ ). Like any other action-reaction pair,  $F_c$  and  $F_F$  do *not* act on the same body, but on two different ones—here on the stone and on the string, respectively, as indicated in Fig. 5-11. Both forces disappear simultaneously when the string breaks.

The physical causes and means for providing a rotating object with the necessary centripetal force vary greatly. A rotating flywheel is held together by the strength of the material itself, which can supply the stress, at least up to a point. A car rounding a curve on a level road is kept from flying off tangentially in a straight line by the force of friction, which supplies the centripetal force by acting (sideways) on the tires. The moon, as Newton first explained, is kept captive in its orbit about the earth by the gravitational pull continually experienced by it. The electron circulating around an atomic nucleus does so by virtue of electric attraction to the center. But in each case, insofar as the motion is circular and uniform,  $F_c = ma_c$ , or, to clothe this expression in all its alternative forms,

$$F_c = ma_c = \frac{mv^2}{r} = m\omega^2 r = m \frac{4\pi^2}{T^2} r = m4\pi^2 n^2 r. \quad (5-8)$$

**PROBLEM 5-12.** Derive, step by step, from first principles, the relation  $F_c = m(4\pi^2/T^2)r$ , and explain at each point what law of physics, concept, approximation, etc. has been introduced.

**PROBLEM 5-13.** The mass of the earth being about  $6 \times 10^{27}$  gm, calculate (by means of the data given in a recent problem) the centripetal force which the sun exerts on it. What is the centrifugal force on the sun?

**PROBLEM 5-14.** The rim of a flywheel 3 m in height is a metal band of 1000 kgm mass. The spokes can support a tension of up to  $10^{11}$  dynes. At what speed of rotation will the wheel burst? Make a drawing to show how the pieces will fly away.

**PROBLEM 5-15.** Mention several other examples of rotational motion, and explain in each case how centripetal force is supplied.

**5-5 Uniformly accelerated circular motion.** In all examples so far, the rotational motion proceeded with constant  $\omega$ . But what if  $\omega$  continually increases or decreases? Let us turn our attention from centripetal accelerations and forces; the only acceleration now to be studied is that along the circular path itself (*tangential* or *linear* acceleration) and the corresponding rate of change of angular veloc-

ity, both produced in general by some force pushing on the whirling object in a direction *tangent* to the circle.

If the motion changes during time interval  $t$  from  $\omega_0$  to  $\omega$ , a new concept suggests itself:  $(\omega - \omega_0)/t$ , the rate of change of angular velocity. This is the *angular acceleration*,  $\alpha$ .

$$\text{By definition,} \quad \alpha = \frac{\omega - \omega_0}{t}. \quad (5-9)$$

Since we shall deal only with cases where  $\alpha$  is constant, Eq. (5-9) may stand indifferently for instantaneous or for average angular acceleration.

It follows from Eq. (5-9) that

$$\alpha = \frac{\omega - \omega_0}{t} = \frac{(v/r) - (v_0/r)}{t} = \left( \frac{v - v_0}{t} \right) \cdot \frac{1}{r} = \frac{a}{r},$$

where  $a$  is the acceleration of the moving point *along* the arc of the circle. Consequently,

$$a = \alpha r. \quad (5-10)$$

Together with two previous relations between linear and angular measure, the complete list is now

$$s = \theta r, \quad v = \omega r, \quad a = \alpha r. \quad (5-11)$$

Having again assumed constant acceleration, we may at once apply to problems of rotation the four equations previously developed (Chapter 1):

$$\left. \begin{aligned} s &= \left( \frac{v + v_0}{2} \right) t, \\ s &= v_0 t + \frac{1}{2} a t^2, \\ v &= v_0 + a t, \\ v^2 &= v_0^2 + 2 a s. \end{aligned} \right\} \quad (5-12)$$

Furthermore, we may obtain another set by substituting the relations in Eq. (5-11); for example, turning to the first of equations (5-12),

$$\theta r = \left( \frac{\omega r + \omega_0 r}{2} \right) t.$$

If carried through for all four parts of Eq. (5-12), we finally obtain the important set of equations for rotational motion with constant  $\alpha$ :

$$\left. \begin{array}{l} \text{I. } \theta = \left( \frac{\omega + \omega_0}{2} \right) t, \\ \text{II. } \theta = \omega_0 t + \frac{1}{2} \alpha t^2, \\ \text{III. } \omega = \omega_0 + \alpha t, \\ \text{IV. } \omega^2 = \omega_0^2 + 2\alpha\theta. \end{array} \right\} \quad (5-13)$$

In practice, the following type of problem might be solved: A flywheel has been running at 2.0 rev/sec, and is now to be sped up to 6.0 rev/sec in 100 complete turns. How long will it take? *Solution:* If  $n_0 = 2 \text{ sec}^{-1}$ ,  $\omega_0 = 2\pi n_0 = 12.6 \text{ rad/sec}$ ;  $\omega = 2\pi (6 \text{ sec}^{-1}) = 37.5 \text{ rad/sec}$ ; for 100 turns,  $\theta = 2\pi \cdot 100 = 628 \text{ radians}$ ; and by Part I of Eq. (5-13),

$$t = \left( \frac{2\theta}{\omega + \omega_0} \right) = \frac{1256}{50.1} \doteq 25 \text{ sec.}$$

**PROBLEM 5-16.** In the previous example, find the angular acceleration. How long will it take the flywheel to coast to a stop from 15 rev/sec with a deceleration of that same value, and how many degrees will a spot on the wheel turn through during that time?

**5-6 Simple harmonic motion.** There is one last type of motion to be investigated here, one closely related to rotation and, as it happens, very frequently observed in nature: vibratory motion. Of this kind is the back-and-forth movement of a pendulum, the up-and-down motion of a bobbing spring, the oscillations of the prongs of a tuning fork and, much more generally, the behavior of all instruments for emitting sound. Vibratory motion is typical of the molecules themselves as sound is propagated through the air, and of all other waves, whether water waves, visible light, radio waves, x-rays or gamma-radiation, or of the electrons in the wires of almost all circuits in use.

Of all the vibratory motions we shall select only the simplest, the special case of Simple Harmonic Motion (abbreviated SHM), to which the examples cited above are often fairly good approximations. The questions we shall ask about bodies executing SHM are of this sort: What is the period or the frequency of vibration? What is the acceleration of the body, and how does it change with time? What is the displacement of a vibrating body (measured from the equilibrium position) at a certain time? All these questions are most easily approached if we first consider a very particular case of SHM, namely,

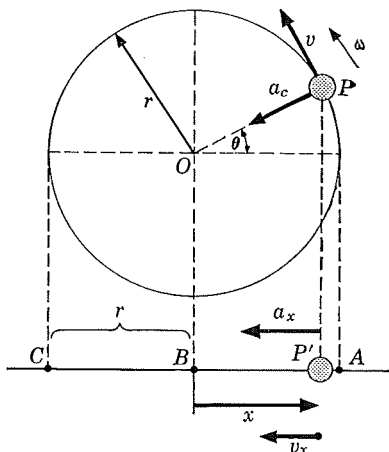


FIG. 5-12. Point  $P$  is in uniform circular motion. Its projection  $P'$  executes simple harmonic motion.

the vibratory motion which is executed by the *shadow projection* of a particle that itself is in uniform circular motion (UCM). For example, Fig. 5-12 shows particle  $P$  going around center  $O$  in a circular path (called the *reference circle*) at constant  $\omega$ . The speed  $v$  is also constant (and has the value  $v = 2\pi r/T$ ), although the direction of  $\mathbf{v}$  changes all the time. As before, the centripetal acceleration needed to maintain UCM is  $a_c = v^2/r = (4\pi^2/T^2)r$ . The symbol  $T$  refers, of course, to the period of rotation, the number of seconds per complete cycle.

Now regard the motion of the *projection* of  $P$ , namely,  $P'$ . You might imagine it to be the shadow cast by the sun shining from directly overhead past a knob  $P$  on a wheel.  $P'$  moves between the limits of  $A$  and  $C$ , back and forth past an equilibrium or average position  $B$ . Its displacement, measured from  $B$  and called  $x$ , is at most  $\overline{BA}$  or  $\overline{CB}$ , two distances that are each equal in length to  $r$ . We say, therefore, that the maximum displacement, which is called *amplitude*, is  $r$ . Next we note that, in general,  $x$  is the projection of line  $\overline{OP}$  at any given moment, therefore

$$x = r \cos \theta, \quad (5-14)$$

where  $\theta$  could also be written as  $\omega t$  or  $(2\pi/T)t$ , i.e.,

$$x = r \cos \left( \frac{2\pi}{T} t \right). \quad (5-15)$$

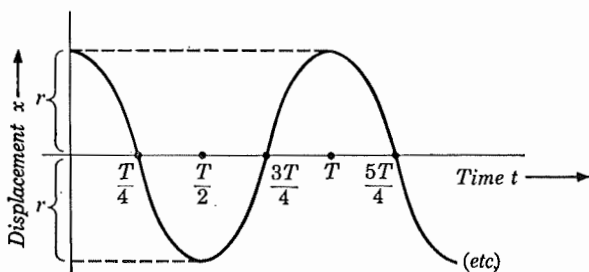


FIG. 5-13. A plot of the equation  $x = r \cos [(2\pi/T) \cdot t]$ .

This may be plotted in graphic form to indicate the displacement of  $P'$  vs. time  $t$  (Fig. 5-13). During one complete period, the displacement goes through a complete cycle, then repeats itself again and again.

Now consider the velocity of  $P'$ . It will simply be the projection ( $v_x$ ), in the horizontal direction, of the instantaneous velocity  $v$  of  $P$ , as shown in Fig. 5-14.

$$v_x = v \sin \theta, \quad (5-16)$$

which may also be written

$$v_x = v \sin \left( \frac{2\pi}{T} t \right) = \left( \frac{2\pi}{T} \right) r \sin \left( \frac{2\pi}{T} t \right). \quad (5-17)$$

Finally, one may inquire about the acceleration  $a_x$  of  $P'$ . It, too, is the projection on the horizontal of the acceleration of  $P$ , in this case the projection of the centripetal acceleration  $a_c$ . As shown in Fig. 5-15,

$$a_x = a \cos \theta, \quad (5-18)$$

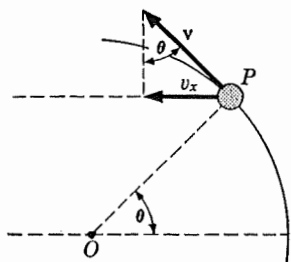


FIG. 5-14.  $v_x = v \sin \theta$ .

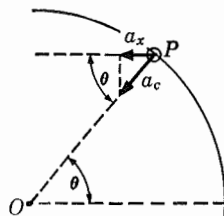


FIG. 5-15.  $a_x = a_c \cos \theta$ .

which we may also write as

$$a_x = a \cos\left(\frac{2\pi}{T}t\right) = \left(\frac{4\pi^2}{T^2}r\right) \cos\left(\frac{2\pi}{T}t\right) \quad (5-19)$$

or, as comparison with Eq. (5-15) reveals,

$$a_x = \frac{4\pi^2}{T^2}x. \quad (5-20)$$

We can summarize all these results in one table but, strictly speaking, one important item has so far been omitted:  $x$ ,  $v_x$ , and  $a_x$  are really vector quantities, and it should be clearly understood that the acceleration vector is at all times directed *toward* the position of equilibrium  $B$ , and is therefore in the opposite direction to  $x$ , the displacement.

Table 5-1. Characteristics of SHM

Value for $P'$ of Position of $P'$	In General	At $A$ or $C$	At $B$
$x$	$r \cos\left(\frac{2\pi}{T}t\right)$	Max, $r$	0
$v_x$	$\frac{2\pi}{T}r \sin\left(\frac{2\pi}{T}t\right)$	0	Max, $\frac{2\pi}{T}r$
$a_x$	$\frac{4\pi^2}{T^2}x$	Max, $\frac{4\pi^2}{T^2}r$	0

From Table 5-1 we may now formulate specifications for SHM, no matter whether executed by the shadow of a body in UCM, or by a spring, or by a pendulum, and so forth. To be truly in SHM, an object or point must move back and forth along a straight line, and must have an acceleration which is proportional to the displacement [see Eq. (5-20)] and directed at all times toward the equilibrium position. Such an object is called a *harmonic oscillator*.

**5-7 The case of the simple pendulum.** As one application of the idea of SHM, let us examine the motion of a simple pendulum, ideally just a point mass ( $m$ ) at the end of a rigid and massless sup-



port of length  $l$  (Fig. 5-16). Our aims are now (a) to see whether the simple pendulum does indeed execute SHM, (b) to derive a formula for the period  $T$  of this pendulum.

(a) The first of these aims is fulfilled if we can show that the path of the bob is (nearly enough) a straight line and that the net force on the bob is proportional to the displacement and always directed to the center (equilibrium) position.

If we restrict the motion of the bob to suitably small angles, the "straight line" condition is automatically met (and  $x \doteq \text{arc}$ ). As to the net force,  $F_{\text{net}}$  in the figure is drawn as the resultant of the tension  $S$  in the weightless string or rod, and of the weight  $W$  of the bob. Note that  $F_{\text{net}}$  must be perpendicular to the string,\* and thus to  $S$ . Therefore, from the drawing,  $F_{\text{net}} = W \sin \gamma$ . But  $\sin \gamma = x/l \therefore F_{\text{net}} = W(x/l)$ , or  $F_{\text{net}}$  is proportional to  $x$  (in magnitude). But this means that the resulting acceleration is also proportional to  $x$ , as required for SHM. The drawing shows also that the force  $F_{\text{net}}$  is always directed to the equilibrium position; hence the bob *will* execute SHM for small angles of swing. For obvious reasons, such a force is called a *restoring force*.

(b) From Eq. (5-20) it followed that the magnitude of the net force on a body in SHM must be

$$F_{\text{net}} = m \cdot a_x = m \frac{4\pi^2}{T^2} x. \quad (5-21)$$

This value for  $F_{\text{net}}$  must correspond to the expression previously calculated, *i.e.*,

$$\frac{m4\pi^2}{T^2} \cdot x = W \frac{x}{l}.$$

Therefore

$$T = 2\pi \sqrt{\frac{ml}{W}}.$$

\* If this point offers difficulty, try to think what would have to happen if  $F_{\text{net}}$  were not  $\perp$  to  $S$ : any component parallel to the string will extend it, move it, or break it.

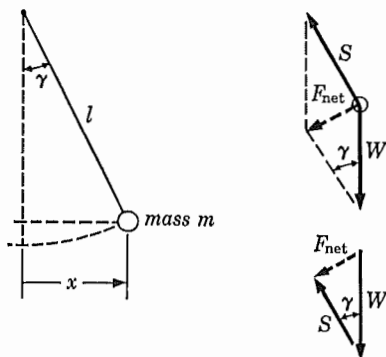


FIG. 5-16. Net force on a pendulum bob.

Note that  $W = mg$ . It follows that the period of a pendulum is given by

$$T = 2\pi\sqrt{\frac{l}{g}}. \quad (5-22)$$

This is a somewhat unexpected result, for it implies that except for air resistance and the like, a heavy and a light pendulum bob will both swing side by side in step if suspended from equally long strings, and that as long as the swing is not too large, say less than six degrees of arc, the period is exactly the same for a medium-sized swing as for a very small one. This is referred to as the *isochronism* of the pendulum, a phenomenon which Galileo is reputed to have discovered experimentally when a young student.

Incidentally, in the equation  $T = 2\pi\sqrt{ml/mg}$ , the two expressions for mass are not conceptually identical:  $m$  in the numerator refers to inertial mass,  $m$  in the denominator to gravitational mass. Only if these two have the same numerical value can we cancel them. Conversely, if  $T = 2\pi\sqrt{l/g}$  is found to hold experimentally, as it does, then we can deduce that these two values are really identical.

Newton himself made pendulum bobs of a great variety of substances, and so found that Eq. (5-22) holds, and that therefore the two types of masses are equivalent. (*Question:* How much more accurate is this confirmation of equivalence than the experiment with the reaction cars?)

**5-8 The vibrating string.** As a further example (Fig. 5-17), let us see whether we can derive the period of oscillation of a light string (length  $L$ , mass  $m$ ) stretched between two rigid supports  $A$  and  $B$  by a tension  $F$  (due to a weight applied to one end, or due to elastic forces in the string). We pull the string upward at the center and then let go. If we neglect dissipation due to friction, etc. the string will then vibrate up and down. Is this SHM, and what is the period  $T$ ?

First we examine the force on the string and make a simplified diagram of it when it is caught in the process of vibration. Selecting in Fig. 5-18(a) a small central portion  $p$  of the string,  $x$  cm above the rest position, note the two forces  $F$  acting on it by virtue of the tension in the string. Evidently, if the string is stretched at all while so displaced, the tension should be larger than at the rest position, but if  $x$  (or  $\epsilon$ ) is very small, we can neglect the difference.

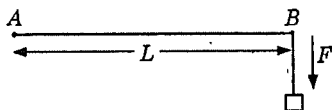


FIG. 5-17.

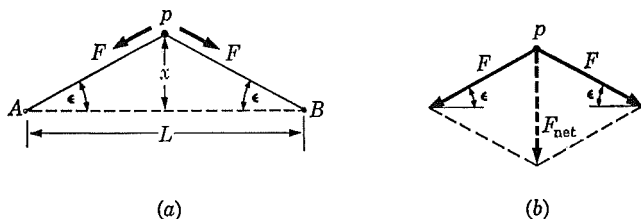


FIG. 5-18. Forces on a segment of vibrating string.

The net force  $F_{\text{net}}$  acting on  $p$  is the resultant of the two vectors  $F$  at equal angles on each side [Fig. 5-18(b)]. By construction,  $F_{\text{net}} = 2(F \sin \epsilon)$ . But  $\sin \epsilon$ , from the figure, is equal to  $\frac{x}{L/2}$  for small angles  $\epsilon$  (where it may be said  $\tan \epsilon \doteq \sin \epsilon$ ). Thus

$$F_{\text{net}} = 4(Fx/L).$$

Since  $F_{\text{net}}$  is proportional to  $x$ , is directed toward a central equilibrium position, and does indeed move  $p$  along a straight line, we expect SHM for the portion  $p$  of the string. Applying Eq. (5-21),

$$4F \frac{x}{L} = m_p \frac{4\pi^2}{T^2} \cdot x, \quad \text{or} \quad T = \pi \sqrt{\frac{m_p L}{F}}.$$

But this equation holds only if the mass  $m_p$  is that of *portion*  $p$ , not the mass of the whole string. On the other hand, the motion of *portion*  $p$  certainly is not unaffected by the rest of the string. Just how to derive an equation for the period  $T$  in terms of the whole string, we now see, is a problem beyond us at the moment.

But we need not give up entirely. One of the reasons for choosing this example is to demonstrate how far one can often go in physical science even when a rigorous solution is not immediately obtainable. Let us ask now: *What kind of expression do we expect for the period of the string?* We have already made many assumptions in this problem which, to a greater or lesser degree, probably would give us only an approximate solution compared with the actual case of a physical vibrating string (e.g., assumptions about the shape of the string and the magnitude and direction of  $F$  on the vibrating portion  $p$ ). In order to get a qualitative solution, we add now another assumption, more obviously and seriously in conflict with rigorous fact but, we hope, not so wrong as to invalidate all our work: the string acts in vibration as though all its mass  $m$  were concentrated at the center portion  $p$ . *To the extent that all these assumptions are right,*

$$T = \pi\sqrt{\frac{mL}{F}}$$

should hold, where  $m$ , the mass of the whole string replaces  $m_p$  of the previous expression.

A more rigorous derivation, however, shows that the correct expression is

$$T = 2\sqrt{\frac{mL}{F}}, \quad (5-23)$$

as is also borne out by experiment. Yet our analysis, while by no means accurate, was not too far off, and it did bring in correctly all the variables that affect the period  $T$ . In this sense we have cause to be pleased with it.

**5-9 Wave propagation.** Let us now consider a specific example of vibratory motion, the propagation of waves. In the previous section we dealt with the vibration of a string fixed at both ends.

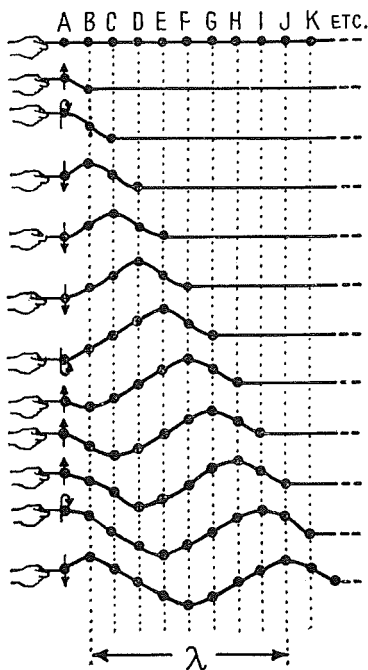


FIG. 5-19. Propagation of a transverse wave along a rope.

Now imagine one that is clamped only at one end. We take hold of the free end, as in Fig. 5-19, and move it rhythmically up and down in simple harmonic motion. As one particle of string is connected with the next, the motion will be communicated from one to the next, but each will lag behind its neighbor's motion, being coupled to it by elastic forces rather than by a rigid connection. We may visualize what happens in terms of the series of successive "snapshots" of Fig. 5-19.

Note how the motion of particle  $A$ , forced upon it by the hand, is communicated to  $B, C, D \dots$  and so on down the line. In particular, the maximum displacement, or *amplitude* of vibration, is reached by each particle in turn. So we see a "crest" being

generated, and after it a "valley"; and as both travel down the string to the right, they are followed by another crest, another valley, and so forth. In short, soon a complete *periodic wave* travels along the string, bringing motion from one end to the other, and with it, of course, energy. Note parenthetically that the wave travels here horizontally while the particles themselves move up and down. Technically, this mutually crosswise motion makes the wave a *transverse* one.

The speed of propagation of the wave  $v_w$ , which we may here take to be the speed with which a crest travels, is of course the ratio of the distance  $s$  which a crest travels on the string to the time  $t$  needed to negotiate that distance. We can evaluate  $v_w$  as follows: If we look at the last of the snapshots in Fig. 5-19, we notice that particles  $B$  and  $J$  are both doing the same thing, i.e., they are both at their maximum displacement upward and about to move down again. The two are "in phase" with each other; and the distance between successive points in phase is called the *wavelength* (symbol  $\lambda$ , the Greek letter lambda). It is clear at a glance how long the crest at  $B$  will take to cover distance  $\lambda$  from  $B$  to  $J$ . Not until particle  $J$  has completed one full cycle of motion down and up will the crest arrive there. The time needed per cycle is, by definition, the observable period of vibration. Therefore,

$$v_w = \frac{\lambda}{T}. \quad (5-24)$$

If we recall that the frequency of vibration  $n$  is the reciprocal of  $T$ , we can write instead

$$v_w = \lambda n, \quad (5-25)$$

a most useful relationship between velocity, wavelength, and frequency, applicable not only to waves in a rope, but to waves of any kind, including light waves. And it is exactly in connection with an examination of the nature of light that we shall again meet these concepts.

**5-10 Conclusion.** Surveying our total progress to this point, we recall that the initial problem was simply to describe uniform motion in a straight line; then followed composite (projectile) motion in a plane. On turning to a consideration of the forces needed to produce such motions, Newton's laws supplied us with concise yet very general answers. Lastly, in this chapter our horizon was extended

still further by the inclusion of the kinetics and some of the dynamics of rotational and vibrational motion. Although there will be occasion to add to this basic store of information on mechanics, we are now ready to study a specific problem of physical science, a *case* as it were, one that lies at the borderline of physics and astronomy, illuminating both while revealing also a little more concerning the origins of modern science.

### Additional Problems

**PROBLEM 5-17.** A clock at Washington, D. C., has a (simple, ideal) pendulum with a period of exactly one second. (a) How long must be the suspension? (Consult Table 4-1.) (b) If the maximum swing is 5 degrees from the vertical, how large is the amplitude of vibration? (c) Plot a graph of the pendulum's displacement *vs.* time as it swings back and forth for two seconds. (d) What will be the position of the pendulum 0.75 sec after the start of a swing? (e) If this pendulum clock were taken to the equator and started there, how much time would it gain or lose in one complete day? (f) Discuss how accurately one might be able to determine the value of  $g$  by means of a pendulum.

**PROBLEM 5-18.** When he wrote the *Principia*, Newton knew that in 1672 Jean Richer had taken a pendulum clock from Paris to Cayenne, French Guiana, to help in astronomical observations; there the clock lost  $2\frac{1}{2}$  minutes each day. Furthermore, from Huygens' work it was then known that the period  $T$  (in seconds) for the pendulum of given length was proportional to  $1/\sqrt{g}$ . If  $g$  for Paris is 980.9, what is  $g$  for Cayenne? (It was this observation which drew general attention to the distinction between the concepts of mass and of weight, until then largely undifferentiated.)

**PROBLEM 5-19.** If a tunnel were to be dug directly to the center of the earth and out at the other side, it would be found that the acceleration of gravity within the tunnel diminishes to zero linearly as the distance to the center decreases. (a) Use this statement to prove that an elevator allowed to drop from the surface down into the tunnel would then oscillate through the earth in SHM. (b) What will be the acceleration of the elevator at the start and at the middle of each round trip? (c) What is the period of oscillation? (d) What is the speed as it passes through the center of the earth? (e) What might be the justification for believing the truth of the initial sentence of this problem?

**PROBLEM 5-20.** Between its two points of suspension a certain violin string is 50 cm long and has a total mass of 2 gm. When played open, the string sounds an A (440 vibration per second). (a) What must be the tension in the string? (b) Where must one put one's finger to play next a C' (528 vib/sec)?

**PROBLEM 5-21.** A fine example of a harmonic oscillator is a vertically suspended (ideally weightless) spring on which hangs a bob of mass  $m$ . When the bob is displaced up or down, the spring is compressed or extended and counters with a restoring force  $F = (\text{a constant } \kappa) \times \text{displacement}$ , where the constant  $\kappa$  depends on the stiffness of the spring. From these data, derive the expression for the period of vibration of a spring with such a bob.

**PROBLEM 5-22.** By measurement, the speed of light (usually symbolized by  $c$ ) is  $c = 2.9979 \times 10^{10}$  cm/sec. The yellow light from the incandescent vapor of sodium has two components, one of wavelength  $5.889953 \times 10^{-5}$  cm, the other  $5.895923 \times 10^{-5}$  cm. Find the corresponding frequency of light of one of the components, but only to an accuracy of three significant figures. (This implies that you may round off each of the data you use to three significant figures before entering upon the calculation. Justify this procedure.)

**PROBLEM 5-23.** Sound waves are longitudinal instead of transverse, i.e., the particles in the path of the wave oscillate in the same direction as the motion of the wave itself. Consequently, the sound wave, instead of crests and valleys, produces successive condensations and rarefactions. Figure 5-20 shows layers of molecules, at first equally spaced, then acted upon by a progressing wave, i.e., by a piston or diaphragm which imparts its own SHM to the gas in front of it. Copy this series and complete it in the same manner as Fig. 5-19 (13 lines altogether). Point out condensations (crowding of the planes) and rarefactions. Then mark off the distance corresponding to one wavelength.

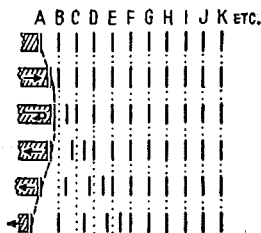


FIG. 5-20. Propagation of a sound wave in air. The displacement of the layers of gas molecules is greatly exaggerated.

### Further Reading

W. F. Magie, *op. cit.* (Ch. 2), pp. 27-30. Huygens' theorem on centripetal force.

L. W. Taylor, *Physics, the Pioneer Science*. Boston: Houghton Mifflin Co. (1941). A splendid text. Chapters 5 and 6 on resolution and composition of vectors.

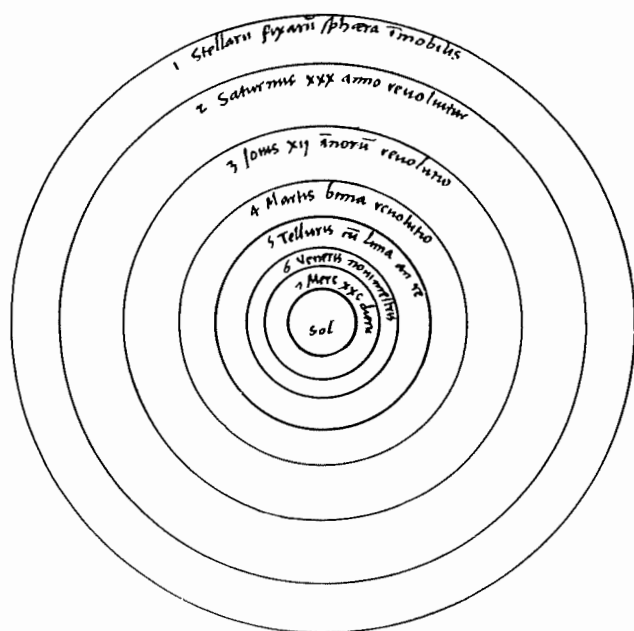




## Part C

# THE LAWS OF PLANETARY MOTION AND UNIVERSAL GRAVITATION

- 6 The Astronomy of Ancient Greece
- 7 Copernicus' Heliocentric Theory
- 8 The Nature of Scientific Theory
- 9 Kepler's Laws
- 10 Galileo's Contribution to Astronomy
- 11 Newton's Law of Universal Gravitation



*The Heliocentric System  
An excerpt from Copernicus' own manuscript.*

## THE LAWS OF PLANETARY MOTION AND UNIVERSAL GRAVITATION

One of man's most persistent and revealing preoccupations has been his attempt to fashion for himself an adequate conceptual model of the Universe. Leaving aside the exciting history of theories concerning the original formation of our solar system—a subject which brings to a focus all that is great in man's intellect, spirit, and poetic imagination—let us confine ourselves to a brief look at the theories concerning the *structure* of our universe as observed by the astronomer.

It is an excellent subject to enlarge upon in a physical science course. The methods of science were perhaps first suspected when man attempted to reduce the apparently chaotic yet recurring motions of the stars, sun, moon, and planets to an orderly system. Second, from early astronomical schemes are derived many of our guiding ideas in science, not only in methodology but also in such specific concepts as time and space, force and motion. We shall see that a continuum of ideas leads from the early naive speculations about stellar motions right to 20th Century science, paralleling the expansion of our universe from the idea of a near sky to the rotating celestial sphere, the recognition of the solar system as a huge galaxy, and finally the universe of innumerable galaxies scattered through unbounded space. Third, the world-system involves many of the men responsible for the strength and content of present physical theory. Their place in science and the atmosphere of their times are all part of our story. Fourth, the work to be discussed—the planetary systems, the law of universal gravitation—had a most profound effect on Western culture. For example, without an understanding of the Newtonian revolution and its background, one cannot hope to comprehend fully the upheavals in 18th Century history, or the development of modern theories of economics and government, or the philosophies of Locke, Berkeley, and Hume. And finally, we may use the opportunity now to watch the rise and fall of physical theories, to analyze their structure, and to sharpen our faculties of discrimination between theories that are useful and those that are misleading. Thus a close examination of this one topic not only will reveal new and significant physical laws, but also will give us a brief chance to inspect those several other aspects of physical science, so important and yet so often inaccessible.

## CHAPTER 6

### THE ASTRONOMY OF ANCIENT GREECE

We can trace the beginning of our development in science to the imaginative minds of the great Greek thinkers. Although the use of optical instruments of precision was still 2000 years away, simple observation of the night sky had by about 400 B.C. established enough data and interpretations concerning the motion of the stars to nourish several world-theories. The "fixed stars" and the Milky Way were commonly thought of as fixed with respect to one another on the surface of a very large sphere, within which was placed the rest of the universe (including the sun, the moon, the earth, and the five easily seen planets). One of the important questions of philosophy at the time was that of the motion of these bodies within the celestial sphere.

Modern scientists might approach such a problem by taking down as many and as accurate data of planetary paths as seem to be needed on the basis of some preliminary guesses or hypotheses, and then might set out to induce a model for the system and the general physical laws of motion for the celestial bodies. These general laws are then accepted as valid if further observations, still more accurate and still more numerous data, bear out the predictions inherent in the laws. But this approach (discussed in detail in Part D) was not fully realized until after the Renaissance. The Greek philosopher, indeed, had invented another approach. On the basis of preliminary observations he formulated a scheme of planetary motion which did not so much need to be in accord with all subsequent observations but instead dovetailed with other, more strictly philosophical or theological problems. The followers of Pythagoras, for example, conceived that the relative sizes of the orbits of planets were proportional to the lengths of the successive strings on a harmoniously tuned stringed instrument. This was thought to assure a "harmony of the spheres," more satisfying and important to the general philosophical requirements of the Pythagoreans than would be the relatively narrower demands of quantitative predictability of physical events made by the modern astronomer. Or, to cite an example we have studied and one involving an entirely different school in Greek science, Aristotle found in qualitative observations of the free fall of large and small bodies enough evidence to satisfactorily integrate these

physical phenomena with his philosophically more wide-reaching, and therefore to him more satisfying, scheme of elements and of natural place.\*

**6-1 Plato's problem.** Nothing is easier and more erroneous than to underestimate the point of view of the early Greeks. Not only did the mechanics of the Greeks "work" within the limits of current interest, but in some schools of thought it was also intellectually a majestic and deeply meaningful construct. Above all, it presents, as it were, the childhood of science, not to be judged in terms of the maturer point of view of contemporary knowledge. It may be true that eventually it outlived its usefulness, but it gave modern men the direction along which they could develop fruitfully. The influence of Greek thought is present in any contemporary activity, in science no less than in art or law, government or education.

Astronomy is a case in point. It is related that Plato (427-347 B.C.) set a problem for his students along these lines: The stars—eternal, divine, unchanging beings—move around the earth, as we can see, in that eminently perfect path, the circle. But a few "stars" appear to wander rather disturbingly across the sky, tracing perplexingly irregular figures in their yearly paths. These are the planets (from the Greek word for *wanderer*). Surely they too must *really* move in uniform, ordered circles, or rather, in their case, combinations of circles. Now, how can we account for the observations on planetary motion and "save the appearances"? Plato's important question may be paraphrased as follows: "Determine what *uniform* and *ordered* motions must be assumed for each of the planets to account for the apparent, more irregular movements." The very formulation of this historic problem strikingly summarizes for us the two main contributions of Greek philosophers to the topic of our discussion:

(a) Physical theory (e.g., the theory of planetary motions) is intelligible only in the context of a prior specific metaphysical theory (e.g., the theory that heavenly bodies must execute "perfect," circular motions).\*\* This proved to be an unfruitful doctrine on which

\* Benjamin Farrington's *Science in Antiquity*, Oxford (1936), is recommended as supplementary reading for these opening paragraphs. Also, consult Chapter 3 of *A Short History of Science and Scientific Thought* by F. S. Taylor, New York: W. W. Norton (1949).

\*\* The term *metaphysics* is used in a specific sense: the discipline which studies principles of knowledge or of being in terms of intuitive, self-evident concepts, concepts of direct "everyday" experience, and analogies.

to base science. By the 16th and 17th centuries, after the great struggle to which we alluded in Chapter 2, it began to be abandoned in favor of the experimental sciences. Whitehead\* makes the following brief analysis of this relatively recent separation of metaphysics from science:

“The reasons for its careful separation from scientific thought are purely practical; namely, because we can agree about science—after due debate—whereas in respect to metaphysics debate has hitherto accentuated disagreement. These characteristics of science and metaphysics were unexpected in the early days of civilized thought. The Greeks thought that metaphysics was easier than physics, and tended to deduce scientific principles from *a priori* conceptions of the nature of things. They were restrained in this disastrous tendency by their vivid naturalism, their delight in first-hand perception. Medieval Europe shared the tendency without the restraint.”

(b) Physical theory is built on observable and measurable phenomena (e.g., the apparent motion of the planets), concerns itself with uniformities of behavior underlying the apparent irregularities, and expresses itself in the language of number and geometry. This guiding idea, which Plato did not extend beyond Astronomy and which was derived in part from the Pythagoreans, was a treasured hint which reappeared when Kepler and Galileo fashioned their experimental science. But Plato held that physical laws can be found from directly intuited principles, the aim being to explain specific phenomena in the context of a philosophic system. The truth of a principle was not measured, as it is today, by its usefulness in every conceivable known or predicted physical situation. Consider this sentence from Plato's *Phaedo*: “This was the method I adopted: I first assumed some principle, which I judged to be the strongest, and then I affirmed as true whatever seemed to agree with this, whether relating to the cause or to anything else; and that which disagreed I regarded as untrue.”

Plato's specific astronomical question, which he did not seriously attempt to answer himself, became the prime concern of astronomers to the time of Galileo. Let us see how they tried to construct a system of the world in accordance with the axiom of “uniform and

---

\* A. N. Whitehead, *The Aims of Education*. New York: Macmillan (1929).

ordered motion," that is, a system which allows the celestial objects to move only in uniform (constant) circular motions or in combinations of several such motions.

**6-2 The Aristotelian system.** Obviously, the first guess to investigate in rationalizing the motion of celestial objects is the most simple geocentric\* system, which places the earth at the center of the celestial sphere carrying the fixed stars. The observed daily motion of those fixed stars will result from our simple model at once if we require the large celestial sphere to rotate uniformly on a north-south axis once a day. We may then attempt to explain the apparent motion of the sun, the moon, and the five visible planets about the fixed earth by letting each be carried on a transparent sphere of its own, one within the other, all seven enclosed by the sphere of fixed stars, and the earth at the center of the whole scheme. But since those seven celestial bodies do not rise and set always at the same point of the horizon, since in fact the planets travel on most complicated paths against the fixed stars, sometimes even briefly reversing the direction of motion, we must give each of their spheres a whole set of simultaneous rotations about different axes, each rotation of the proper speed and direction, each axis of the proper inclination, to duplicate the actually observed paths.

Here the mathematical genius of the Greek was indeed put to the test! The sun and the moon were relatively simple to discuss, but some of the planets offered much difficulty. Plato's pupil Eudoxus thought that 26 simultaneous uniform motions would do for all seven celestial bodies. In Aristotle's work are prescribed 29 additional uniform motions, partly to account for some of the most obvious discrepancies between Eudoxus' system and the observed paths.\*\* Figure 6-1 represents in barest outline the most important features of a medieval Aristotelian scheme of the Universe. Only the main spheres are indicated, the multitude of mutually supporting axes, etc., is not shown.

And yet, there remained easily observed features of the sky which were unexplained by Aristotle's system—notably the fact that the sun, the moon, Venus, Mars, and Jupiter at times seemed nearer and at other times farther away from the earth. A set of uniform rota-

---

\* Earth-centered, from the Greek *gē*, earth.

\*\* See *A Source Book in Greek Science* by Morris R. Cohen and I. E. Drabkin [New York: McGraw-Hill (1948)] for most illuminating translations from the original sources.

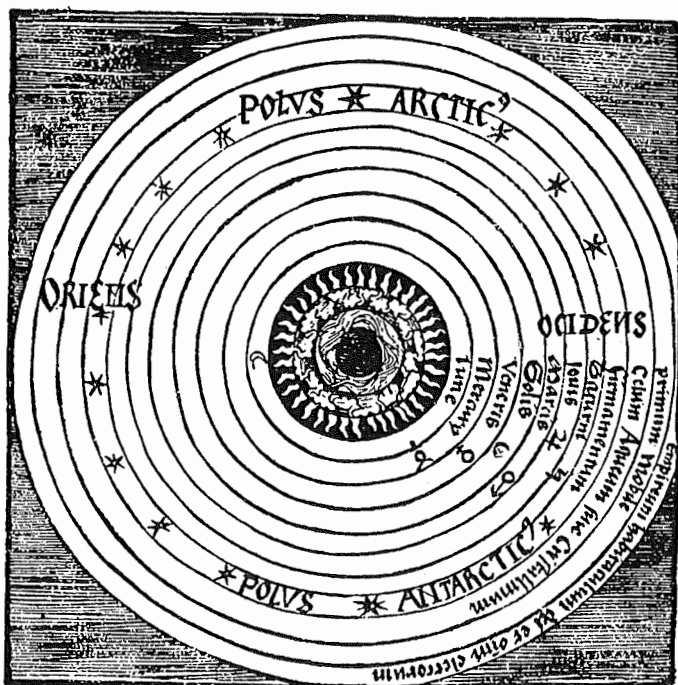


FIG. 6-1. A medieval conception of the world. The sphere of the moon (*luna*) separates the terrestrial region (composed of concentric shells of the four elements Earth, Water, Air, and Fire) from the celestial region. Beyond the moon are the concentric spheres carrying Mercury, Venus, Sun, Mars, Jupiter, Saturn, and the fixed stars. (From a wood cut of 1508.)

tions of the celestial bodies on spheres *concentric* with the earth could never allow them to change their distance from the earth. Aristotle was aware of this, but discounted the importance of the simple, but eventually fatal, argument against the fundamental hypothesis of his system. He did not take the consequences which to us seem deceptively obvious, mainly because it was then, as it is now, incomparably easier to leave in doubt the testimony of a few contrary observations than to give up a whole world scheme which seemed conclusive and necessary from the point of view both of his philosophy and his mechanics. Of course, this does not mean that he proposed a theory he knew to be false. Remember the discussion in the chapter on the motion of free fall, where it was emphasized that the original Aristotelian science was not simply bad modern science but an activity fundamentally different from it. Perhaps we may also

see here an example of an important human trait which colors all scientific work and which even the greatest thinkers can never hope to overcome entirely—we all tend to deny the importance of facts or observations not in accord with our convictions and preconceptions, so that sometimes we ignore them altogether, even though, from another point of view, they stand before our very eyes. Moreover, even the most general and modern scientific theory does not hope or even seriously attempt to accommodate every single detail of every specific case. One is always forced to idealize the observations before attempting a match between “facts” and theory—not only because there are usually unavoidable experimental uncertainties in observation, but because conceptual schemes are consciously designed to apply to *selected* observations rather than to the totality of raw experience. As a consequence, the history of science is studded with cases in which it turned out too late that the neglected part of a phenomenon was actually its most significant aspect. But on the other hand, if we were to allow no tentative, half-true, or plainly wrong theories in science, we should probably never see a correct one evolving at all. Since the task cannot humanly be done in one jump, we must be satisfied with successive approximations.

**6-3 The heliocentric\* theory.** Of course, the problem of planetary motion persisted. There were two different major types of attack after Aristotle—the heliocentric theory and the modified geocentric theory. Let us now discuss the first of these. Aristarchus of Samos (third century B.C.), perhaps influenced by the work of Heraclides of Pontus (fourth century B.C.), suggested that a simple world system would result if the sun were put at the center of the universe and if the moon, the earth, and the five then known planets revolved around the sun in orbits of different sizes and speeds. We do not know many details; his work on this subject is known to us only through references in other old writings. But evidently he assumed that the earth has a daily rotation on its north-south axis, as well as a yearly revolution in the orbit around the sun, and he placed the whole system in the sphere of fixed stars, which thereby could be considered at rest with respect to the center of the universe.

This heliocentric hypothesis has one immediate advantage. It explains the bothersome observation that the planets are at times nearer to the earth and at other times farther away. But the ancient world saw three very serious flaws in Aristarchus' suggestion:

---

\* From the Greek *helios*, sun.



*First*, it did violence to philosophical doctrines (e.g., that the earth, by its very "immobility" and position is differentiated from the "celestial bodies," and that the natural "place" of the earth is the center of the universe). In fact, his contemporaries considered Aristarchus impious "for putting in motion the hearth of the Universe." Also, the new picture of the solar system contradicted common sense and everyday observation; the very words used in astronomy (sunrise, progression of planets, etc.) reflect the intuitive certainty that the earth must be at rest.

PROBLEM 6-1. List the "common-sense" observations concerning sun, stars, and planets from which the geocentric theory sprang, apparently so convincingly. Be careful not to *assume* the geocentric or any other theory in your description.

*Second*, Aristarchus does not seem to have fortified his system with detailed calculations and quantitative predictions of planetary paths, by our present standards an obvious condition for assuring recognition in the physical sciences. This work seems to have been purely qualitative, although in some other accomplishments he is said to have showed considerable mathematical powers.

*Third*, the Greek thinkers offered an ingenious bit of reasoning to refute Aristarchus. If the earth is to move around the sun, its large orbit will carry it sometimes near to a given fixed star on the celestial sphere, and sometimes farther away from it. Thus the angle at which we have to look for this star will be different as seen from the various points in the earth's annual sweep. This phenomenon, called the annual *parallax* of the fixed stars, should occur on the basis of Aristarchus' heliocentric hypothesis, but it was not observed by the Greek astronomers.

In explanation we may say either (a) that the stellar parallax is so small as to be unobservable with the naked eye—which in turn requires the fixed stars to be unimaginably distant compared with the diameter of the yearly orbit of the earth, or (b) that Aristarchus was wrong and the earth does not move around within the celestial sphere. It seems natural that the ancients, predisposed to reject the heliocentric system, and also loath to consider an infinitely extensive universe, chose the second of these alternatives. The first, however, proved eventually to be the correct one. The parallax was indeed present, although so small that even telescopic measurements did not reveal it until 1838.

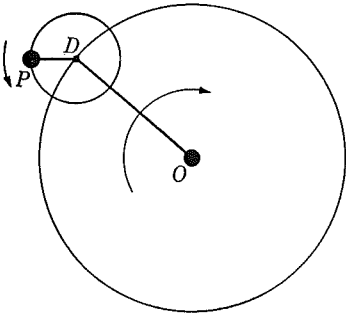


FIG. 6-2. An example of epicyclic motion of a planet  $P$ .

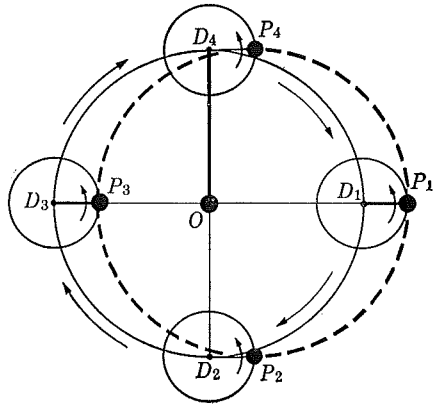


FIG. 6-3. Eccentric path represented by epicyclic motion.

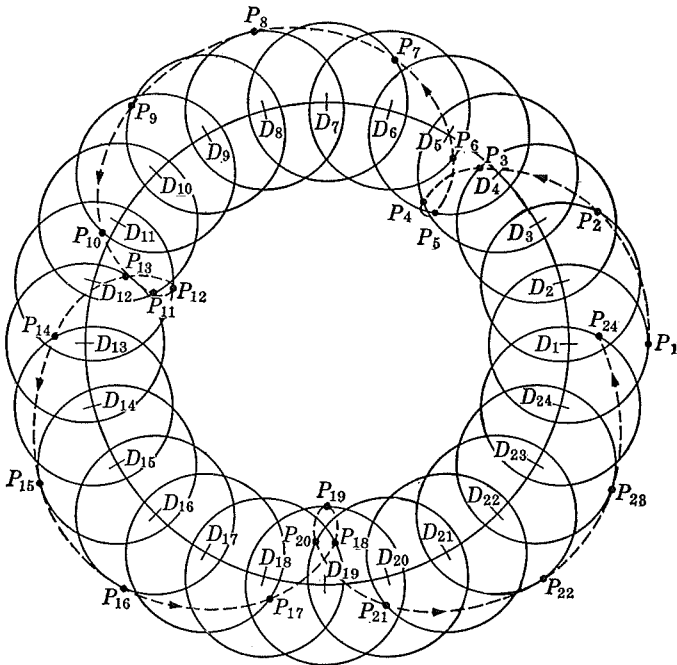


FIG. 6-4. Epicyclic motion of  $P$ , showing three temporary reversals of direction of motion at  $P_4, P_{11},$  and  $P_{18}$ . (Adapted from M. Cohen and I. Drabkin.)

**PROBLEM 6-2.** The annual parallax of a given star may be crudely defined as half the angle between the two lines of sight drawn from the center of the earth to the star from opposite ends of a diameter across the earth's orbit. Since the Greeks could not detect the parallax, and since their accuracy of measurement was often only about  $\frac{1}{2}^\circ$ , what must be the least distance from the earth's orbit to the nearest fixed star? (Express in astronomical units, or A.U., one A.U. being the sun's mean distance from us, or about  $93 \times 10^6$  miles.) F. W. Bessel in 1838 first observed the annual parallax, which for the nearest bright star,  $\alpha$  Centauri, is of the order of  $\frac{3}{4}$  second of arc. What is its approximate distance from us? (Modern telescopes can measure parallaxes of about  $0.01''$ , and therefore can directly determine distances of stars about 100 times as far away as  $\alpha$  Centauri.)

The heliocentric theories of Aristarchus and others seem to have been so uninfluential in Greek thinking that we should ordinarily spend no time on them. But these speculations stimulated the crucial work of Copernicus 18 centuries later. Fruitful ideas, it is evident, are not bound by time or space, and can never be evaluated with *final* certainty.

**6-4 Modified geocentric theories.** We now turn to the other and more vigorous offshoot of the early astronomical schemes. To allow the planets to have variable distances from the earth while still keeping to the old belief of an immovable earth, the system of concentric spheres was modified in ingenious ways, principally by Hipparchus of Rhodes (second century B.C.) and the astronomer and geographer, Claudius Ptolemy of Alexandria (second century A.D.).

(a) *Eccentric motion.* If the stationary earth were not exactly at the center of rotation of a uniformly moving celestial object, the latter would move along an *eccentric* path as seen from the earth, and vary in distance at different times. This scheme fits the apparent yearly motion of the sun fairly well, since he appears larger (and consequently nearer) at noon in our winter compared with our summer. Note that by admitting eccentric motions the astronomers really were dodging the old principle that required the planet motions to be circular around the center of the earth.

(b) *Epicyclic motion.* Figure 6-2 represents an object  $P$  (such as the sun or a planet) having two simultaneous uniform rotary motions, one a rotation of  $P$  about a point  $D$  (in space) with radius  $PD$ , the other a rotation of the line  $OD$  about point  $O$  (the position of the earth). The small circle is called an epicycle, the large circle the deferent. The two motions may have entirely independent speeds,

directions, and radii. Figure 6-3 indicates the special case where the two combined motions yield an eccentric path (dotted line), and Fig. 6-4 shows, by the dotted line connecting 24 successive positions of  $P$ , the complicated motion resulting when  $P$  rotates about  $D$  several times while  $D$  moves once around  $O$ .

**PROBLEM 6-3.** What is the ratio of the speeds of revolution along the epicycle and the deferent in Fig. 6-4? What might be the path if the ratio were exactly 3:1?

This last type of motion does, in fact, exhibit most of the observed complications in the paths of planets as seen from the earth against the background of the fixed stars; note particularly the reversal in the direction of motion ("retrograde motion") at positions  $P_4$ ,  $P_{11}$ , and  $P_{18}$ . Jupiter's path, one with 11 such loops, is covered in approximately 12 years per complete revolution around the earth.

By a proper choice of radii, speeds, and directions, an eccentric path can be traced just as well by an epicyclic motion. Ptolemy used one or the other, and sometimes mixtures of both types, depending on the problem. We must also note that the device of epicyclic motion suffers from the same philosophical difficulties as the eccentric. The rotation of  $P$  about  $D$  is, if we wish to be strictly logical, just as much in conflict with old axioms as was the circular motion of  $P$  about a center other than the earth.

**PROBLEM 6-4.** Construct as well as you can with compass and ruler the path of  $P$  (similar to Fig. 6-3), if  $P$  revolves around  $D$  twice while  $D$  goes around  $O$  once.

**PROBLEM 6-5.** From observations we may form the following hypotheses. The sun moves around the earth from east to west once a day, just like the celestial sphere, but about four minutes per day less quickly. The sun also approaches the earth during our winter and recedes in summer, and in addition it slowly travels north of the equator in summer and south of the equator in winter. Suggest and draw the outlines of a qualitative geocentric model for these motions of the sun and celestial sphere about the earth, using some of the devices discussed.

Now we can summarize what has been retained from Plato and Aristotle, and what has been amended. Use is still made of *uniform and circular motions*, and of a *stationary earth*. Vanished is the scheme of spheres all concentric at the earth, and with it the need to have

all rotations exactly earth-centered. This is even more strikingly evident when we discover that Ptolemy found it necessary to add still another device to the list in order to represent more faithfully some features of celestial motion. This device is the *equant*.

(c) *The equant.* Figure 6-5 shows an object  $P$  in cyclic motion around  $D$ , which in turn moves on a circle whose center is at  $O$ . The earth may be at  $O$  or, if this is a mixture of epicyclic and eccentric motion, the earth may be anywhere along the line  $AA^1$ , say at position  $E$ . So far the motion of  $D$  has been specified as uniform with respect to  $O$ , but to represent some planetary motion in Ptolemy's system it was necessary to let  $D$  revolve uniformly with respect to  $Q$ , the equant. That is, the angle  $DQA$  changes at a constant rate while  $D$  executes its circular sweep. Now  $D$  is no longer in strictly uniform circular motion, although its motion is still *uniform* (seen from  $Q$ ) and *circular* (seen from  $O$ ).

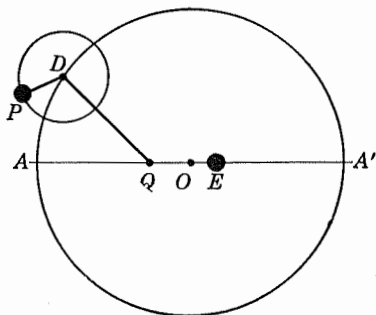


FIG. 6-5. Motion with respect to equant at  $Q$ .

Ptolemy's great work, known through Arab translations as the *Almagest*, built up a system of the world on the basis of these devices and the following preliminary assumptions:

- "(1) that the heaven is spherical in form and rotates as a sphere;
- (2) that the earth, too, viewed as a complete whole, is spherical in that form;
- (3) that it is situated in the middle of the whole heaven, like a center;
- (4) that by reason of its size and its distance from the sphere of fixed stars the earth bears to this sphere the relation of a point;
- (5) that the earth does not participate in any locomotion."

Implicit in his work is also the old, now somewhat distorted doctrine of uniform circular motions as the only behavior thinkable for celestial objects.

PROBLEM 6-6. What is the significance of the fourth point in Ptolemy's preliminary assumptions?

**6-5 The success of the Ptolemaic system.** By adjusting the respective axes, directions of motions, rates and radii of rotations, number and size of epicycles, eccentrics, and equants, by fitting his devices to the observed paths in a cut-and-try fashion, by utilizing partial results of generations of previous astronomers—so did Ptolemy manage to assemble the apparatus which proved still useful to astronomers, navigators, and astrologers more than 14 centuries later. It was an answer to Plato's original question, and on the whole a magnificent piece of work. One pleasing feature was that centers of the epicycles of the moon, Mercury, and Venus fell on the same line between Earth and Sun (see Fig. 6-6).

But the complexity and inaccuracy of detail in the whole scheme was considerable, and subsequent observations required amendment of some features of the model from time to time, e.g., the change of an assigned equant or the addition of another epicycle. By the time of Copernicus, this geocentric system required more than 70 simultaneous motions for the seven celestial bodies. But against this were five powerful reasons why the system was generally accepted wherever it became known:

(a) It gave an accurate enough description of what could be observed with the instruments of the times.

(b) It predicted the future paths of planets well enough for the purposes of the times, if only after cumbersome calculations; and

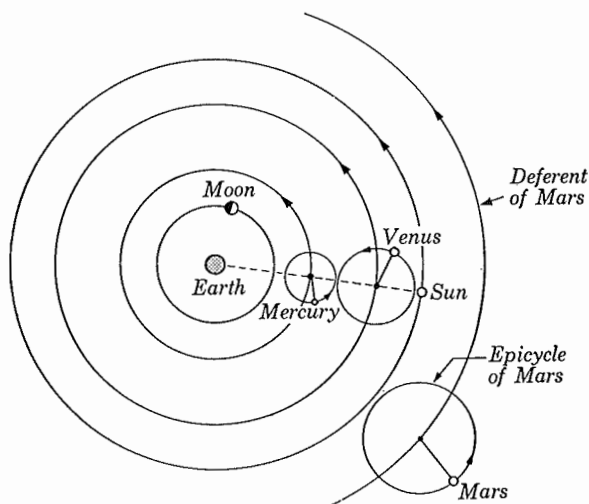


FIG. 6-6. Partial and schematic diagram of Ptolemaic system of planetary motion.

when a serious discrepancy between predictions and observations occurred it could be resolved by tampering a little with the gears of the flexible apparatus. Today, even as at that time, the geocentric system is still preferred for calculations in navigation—and in astrology!

(c) It explains naturally why the fixed stars show no annual parallax.

(d) In most details it coincided with Greek philosophical and physical doctrine concerning the nature of the earth and the celestial bodies. Later, when reintroduced to Europe by the Arabs, the Ptolemaic system\* was given theological significance by the scholastics. Furthermore, it was also in line with contemporary physics (e.g., projectile motion), based on the same philosophical doctrine of natural motion, natural place, and so forth. (See Chapter 2.)

(e) It had "common-sense appeal." It is almost unavoidable to hold the opinion that we actually can "see" the sun and stars moving around us, and it is comforting to think of ourselves on a stable, immovable earth.

But for all that, Ptolemy's theory was eventually displaced by a heliocentric one. Why did this happen? What were the most significant deficiencies in Ptolemy's picture? When is a scientific theory, from our present point of view, successful or unsuccessful? We shall want to answer these questions in detail after we have looked at an outline of the great rival scheme of planetary motion.

---

\* In connection with our study of the Ptolemaic system there is surely no more thrilling experience than to read that immortal excursion through the medieval universe, the *Paradiso* of Dante (for example, in the 1944 edition of the *Modern Library*).

### Further Reading

M. R. Cohen and I. E. Drabkin, *A Source Book in Greek Science*. New York: McGraw-Hill (1948). A wealth of source material and penetrating comments. See pp. 90-143 for Astronomy.

B. Farrington, *Science in Antiquity*. London: Oxford University Press (1936, 1947). Home University Library.

B. Farrington, *Greek Science*. Pelican Books A142 and A192 (1944, 1949).

Sir Thomas Heath, *The Copernicus of Antiquity (Aristarchus of Samos)*. London: Society for Promoting Christian Knowledge (1920).

Sir Thomas Heath, *Greek Astronomy*. London: J. M. Dent (1932).

C. Singer, *A Short History of Science*. Oxford: Clarendon Press (1941). Very useful.

F. S. Taylor, *A Short History of Science and of Scientific Thought*. New York: W. W. Norton (1949), Chapter III.

W. P. D. Wightman, *The Growth of Scientific Ideas*. New Haven: Yale University Press (1951), Chapters I-IV.



## CHAPTER 7

### COPERNICUS' HELIOCENTRIC THEORY

**7-1 Europe reborn.** Our study takes us now to Renaissance Europe of the years around 1500 A.D. Astronomical theory had not progressed significantly since Ptolemy. St. Thomas Aquinas (1225-1274) had joined the Aristotelian ideas of celestial motions with Christian theology. The geocentric theory had received new meaning in terms of current philosophical doctrine; to question one was to attack the other, and so far there had appeared no one who thought it necessary or had the daring to battle seriously against such formidable allies.

But the "Dark Ages" were passing; the Renaissance movements spread out from Italy to sweep the Western world, and within a few generations there arose a new Ideal of Man, the confident individual full of curiosity and *joie de vivre*. (You may strikingly achieve a feeling for the changing atmosphere by comparing extremes, perhaps the paintings of a Hieronymus Bosch on the one hand with those of a Botticelli on the other.) In the topography of world history this great period of summing up old wisdom and of forming new attitudes is analogous to a watershed on which great rivers have their sources. It was the time in which lived as contemporaries or within a generation of one another most of the men whose work heralded the new age: Gutenberg and da Vinci, Columbus and Vasco da Gama, Michelangelo and Dürer, Erasmus, Vesalius and Agricola, Luther and Henry VIII. One cannot help comparing this short time interval of decisive change with a similar break in the traditions of Western thought four hundred years later—a period which, between the publication of Darwin's *The Origin of Species* (1859) and the first large-scale release of atomic energy, bracketed such names as Mendel and Pasteur, Planck and Einstein, Rutherford and Bohr, and also such new forces as Marx and Lenin, Freud and Pareto, Picasso and Stravinsky, Shaw and Joyce.

**7-2 The Copernican system.** When Nicolaus Copernicus (1473-1543) was a young student in Poland, the New World was discovered. During his exciting life he watched in his old world the gathering of great cultural changes. And on the very day he died he saw the



FIG. 7-1. Nicolaus Copernicus (1473-1543). (From a woodcut of about 1590.)

first copy of his great book, the *Revolutions*, which gave us a whole new universe.\*

The full title of Copernicus' main work, *Six Books Concerning the Revolutions of the Heavenly Spheres*,\*\* startles us at the very outset with the implication of an Aristotelian notion of concentric spheres. He was indeed concerned with the old problem of Plato, the construction of a planetary system by combination of the fewest possible

\* A highly readable and stimulating presentation is *Sun, Stand Thou Still; The Life and Work of Copernicus the Astronomer*. New York: Henry Schuman Inc. (1947); also in paper-bound pocket edition. The account reveals to us a most remarkable astronomer, mathematician, churchman, administrator, diplomat, physician, and able student of the classics and of economics—and yet a humble and compassionate man.

\*\* Excerpts will be found in Knedler's and many other collections. The full dedication is printed as Appendix C in W. T. Sedgwick and H. W. Tyler's *A Short History of Science*. New York: Macmillan (1929).

uniform circular motions. Fundamentally, Copernicus (see Fig. 7-1) was motivated to postulate his heliocentric system because it avoided two assumptions inherent in the geocentric theory as taught at the time, to which he took particular exception: (a) that the notion of the equant could be used freely, and (b) that the earth was to be considered at rest and at the center of the universe as a matter of dogma, which greatly complicated the description of celestial motions. But let us hear Copernicus' own arguments on these two points:

(a) “. . . the planetary theories of Ptolemy and most other astronomers, although consistent with the numerical data, seemed . . . to present no small difficulty. For these theories were not adequate unless certain equants were also conceived; it then appeared that a planet moved with uniform velocity neither on its deferent nor about the center of its epicycle. Hence a system of this sort seemed neither sufficiently absolute nor sufficiently pleasing to the mind.

“Having become aware of these defects, I often considered whether there could perhaps be found a more reasonable arrangement of circles, from which every apparent inequality would be derived and in which everything would move uniformly about its proper center, as the rule of absolute motion requires.”

To Copernicus, any type of celestial motion other than uniform circular motion was “obviously” impossible: “. . . the intellect recoils with horror” from any other suggestion; “. . . it would be unworthy to suppose such a thing in a Creation constituted in the best possible way.” These arguments are of the same type as those of his scholastic opponents, except that to them the immobility of the earth was equally “obvious.”

(b) Reading the classics, Copernicus found:

“. . . according to Cicero, Nicetas had thought the earth moved, . . . according to Plutarch certain others [including Aristarchus] had held the same opinion . . . when from this, therefore, I had conceived its possibility, I myself also began to meditate upon the mobility of the earth. And although it seemed an absurd opinion, yet, because I knew that others before me had been granted the liberty of supposing whatever circles they chose in order to demonstrate the observations concerning the celestial bodies, I considered that I too might well be allowed to try whether

sounder demonstrations of the revolutions of the heavenly orbs might be discovered by supposing some motion of the earth. . . . I found after much and long observation, that if the motions of the other planets were added to the motions [daily rotation, and yearly revolution about the sun] of the earth, . . . not only did the apparent behavior of the others follow from this, but the system so connects the orders and sizes of the planets and their orbits, and of the whole heaven, that no single feature can be altered without confusion among the other parts and in all the Universe. For this reason, therefore, . . . have I followed this system."

Thus, in essence, Copernicus proposed a change of point of view for the representation of celestial motions along the lines of Aristarchus' heliocentric system. By highly gifted calculations, using some observations of his own, Copernicus then proved what Aristarchus had failed to do: that the motion of celestial objects as then known could indeed be represented by a combination of a few uniform circular motions in a sun-centered system. All planets, the earth included, could be imagined as moving on concentric spheres, with only relatively few and small epicycles and eccentrics still needed to account for the finer details of motion.\* Moreover, the *same direction* of motion along almost all deferents and epicycles could be assumed, which had not been true for the geocentric model. But above all, the odious equant could be discarded. All motions were now truly circular and uniform with respect to their own centers. Plato's question was answered again, in an alternative way.

Copernicus' diagram (Fig. 7-2) shows the main concentric spheres carrying the planets around the sun. His accompanying text explains the outlines of the system and, incidentally, gives us a magnificent insight into his deeper motivations:

"The ideas here stated are difficult, even almost impossible, to accept; they are quite contrary to popular notions. Yet with the help of God, we will make everything as clear as day in what follows, at least for those who are not ignorant of mathematics. . . .

"The first and highest of all the spheres is the sphere of the fixed stars. It encloses all the other spheres and is itself self-contained; it is immobile; it is certainly the portion of the universe

---

\* The common center of the main spheres actually had to fall a little to one side of the unmoving sun, so that this system is not *quite* heliocentric, just as Ptolemy's earth-position did not coincide with the center of the sun's motion.

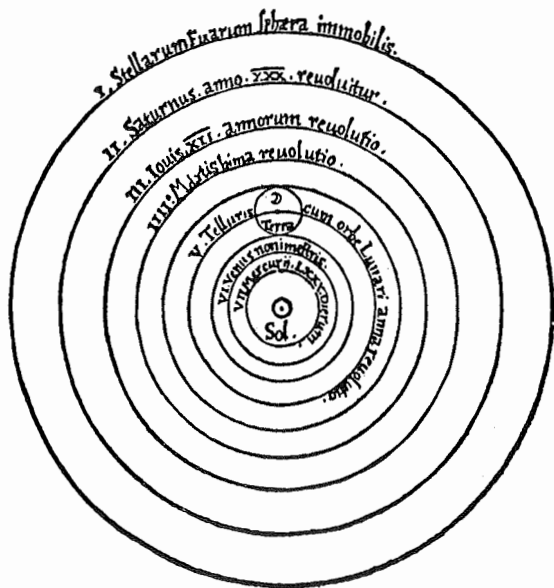


FIG. 7-2. Copernicus' diagram of his Heliocentric System. (From his own work, 1543.)

with reference to which the movement and positions of all the other heavenly bodies must be considered. If some people are yet of the opinion that this sphere moves, we are of a contrary mind; and after deducing the motion of the earth, we shall show why we so conclude. Saturn, first of the planets, which accomplishes its revolution in thirty years, is nearest to the first sphere. Jupiter, making its revolution in twelve years, is next. Then comes Mars, revolving once in two years. The fourth place in the series is occupied by the sphere which contains the earth and the sphere of the moon, and which performs an annual revolution. The fifth place is that of Venus, revolving in nine months. Finally, the sixth place is occupied by Mercury, revolving in eighty days.

"In the midst of all, the sun reposes, unmoving. Who, indeed, in this most beautiful temple would place the light-giver in any other part than that whence it can illumine all other parts? . . .

"In this orderly arrangement there appears a wonderful symmetry in the universe and a precise relation between the motions and sizes of the orbs which is impossible to obtain in any other way."

**7-3 Bracing the system.** Knowing well that to many his work would seem absurd, "nay, almost contrary to ordinary human understanding," Copernicus attempted to fortify it against anticipated criticism in four ways.

(a) He tried to make plausible that his assumptions coincided with dogma at least as well as Ptolemy's. He has many passages (we have seen samples) on the deficiencies of the Ptolemaic system, on how harmonious and orderly his own seems, and on how pleasingly and evidently it reflects the mind of the Divine Architect. To Copernicus, as to all men of his period (and to many of our own), the observable world was but a symbol of the working of God's mind; to find symmetry and order in the apparent chaos of sense data was to him an act of reverence, a renewed proof of the Deity. He would have been horrified if he had known that his theory was ultimately responsible for the sharp clash, in Galileo's time, between science and dogma. Copernicus, we must not forget, was a highly placed and honored church dignitary. In matters of doctrine he might have regarded himself as a conservative who was objecting to current scholastic teaching only to the extent of wishing to make it more consistent with Aristotle's principles of mechanics. (Compare the forms of the systems in Figs. 6-1 and 7-2.)

(b) Copernicus prepared enough quantitative material to put his book mathematically on the same footing as Ptolemy's, that is, he calculated the relative radii and speeds of the elements in his system so that tables of planetary motion could be made. (Note how this compares with Aristarchus' more qualitative work.) The two theories were thus equally correct with respect to prediction of future planet positions within the then current error of observation of at least  $\frac{1}{3}$  degree of arc. However, on looking at Copernicus' treatise, we must not expect to find a mathematical treatment of the type seen in modern texts. Remember that even our simplest trick of mathematical notation, + and - signs, did not come into use until after Copernicus' death.

(c) With considerable ingenuity and much success, Copernicus tried to answer several objections that were as certain to be raised against his heliocentric system as they had been, long ago, against Aristarchus'. To the argument that his earth, rotating so rapidly about its own axis, would surely burst like a flywheel driven too fast, he countered, "Why does the defender of the geocentric theory not fear the same fate for his rotating celestial sphere—so much faster because so much larger?" To the argument that birds in flight and

the like should be left behind by the rapidly rotating and revolving earth,\* he answered that the atmosphere is dragged along with the earth. As to the old question of the absence of parallax among the fixed stars, he could give only the same answer as Aristarchus (a good answer even though the Greeks did not accept it):

“. . . the dimensions of the world are so vast that though the distance from the sun to the earth appears very large compared with the size of the orbs of some planets, yet compared with the dimensions of the sphere of fixed stars, it is as nothing.”

This distance to the fixed stars, he says elsewhere, is “so immense as to render imperceptible to us even their apparent annual motion. . . .”

He cites several other good arguments (and a few erroneous ones) to support his system, but these will sufficiently illustrate his point.

(d) To us the most impressive success of Copernicus is the reduction in the number of elements needed in his system (to about 34) and the consequent greater ease with which his apparatus can be used by the astronomer for the solution of practical problems. Science has learned to cherish these features: economy of concepts and assumptions, simplicity of formulation, applicability to a variety of problems. Copernicus knew well the advantage he held over the rival theory, and he hoped that he would be accepted on this basis, irrespective of current prejudices.

**PROBLEM 7-1.** Make a list of assumptions underlying Copernicus' theory, analogous to the list for Ptolemy's (see Section 6-4).

**7-4 The opposition to Copernicus' theory.** But Copernicus' hope was not quickly fulfilled. It took more than 100 years for the heliocentric system to be generally accepted by astronomers; in the meantime the theory and its few champions were faced with powerful opposition, some of it the very same which antagonists had used against the heliocentric ideas of the Greek philosophers.

(a) First came the argument from dogma concerning the immobility and central position of the earth. For all his efforts, Copernicus was in general unsuccessful in persuading his readers that the heliocentric system was at least as close as the geocentric one to the mind and intent of the Deity. All religious faiths in Europe, including

---

\* The orbital speed around the sun is about 70,000 mi/hr, and the speed of west-east rotation alone of a point on the equator is more than 1,000 mi/hr.

the newly emerged Protestants, found enough Biblical quotations (e.g., Joshua 10:12-13) to assert that the Divine Architect had worked from a Ptolemaic blueprint.

PROBLEM 7-2. Read Joshua 10:12-13 and interpret the astronomical event referred to in terms of the two rival theories.

Martin Luther himself branded Copernicus a fool and heretic. The Papacy eventually put the *Revolutions* on the Index of forbidden books as "false and altogether opposed to Holy Scriptures," withdrawing its approval of an earlier outline of Copernicus' work. Some Jewish communities forbade the teaching of the heliocentric theory. It was as if man's egocentric philosophy demanded the middle of the stage for his earth, the scene both of his daily life and prayer in a world created especially for his use, and of the drama of salvation with the expected eventual coming of the Savior or Messiah.

Although it was allowed that for mathematical simplicity the Copernican scheme might indeed be useful, and although even St. Thomas Aquinas had grave misgivings about the Ptolemaic system because it did not stick to strictly uniform circular motions as interpreted by Aristotle, yet it seemed "philosophically false and absurd," dangerous, and fantastic to abandon the geocentric hypothesis. And in all truth, what other reaction could one have expected? We must try to appreciate the frame of reference of the time. In general, Europe then recognized as its two supreme sources of authority the Bible and Aristotle. Physical science as we now think of it was only gradually evolving; that was still the great period of learned and respected Astrologers (like Nostradamus) and Alchemists (like Paracelsus).

(b) A further explanation for the resistance which the Copernican theory had to overcome is to be sought in the damage the Copernican theory brought even to contemporary *physics*. This is well discussed by H. Butterfield:

". . . at least some of the economy of the Copernican system is rather an optical illusion of more recent centuries. We nowadays may say that it requires smaller effort to move the earth round upon its axis than to swing the whole universe in a twenty-four hour revolution about the earth; but in the Aristotelian physics it required something colossal to shift the heavy and sluggish earth, while all the skies were made of a subtle substance that was supposed to have no weight, and they were comparatively



easy to turn, since turning was concordant with their nature. Above all, if you grant Copernicus a certain advantage in respect of geometrical simplicity, the sacrifice that had to be made for the sake of this was nothing less than tremendous. You lost the whole cosmology associated with Aristotelianism—the whole intricately dovetailed system in which the nobility of the various elements and the hierarchical arrangement of these had been so beautifully interlocked. In fact, you had to throw overboard the very framework of existing science, and it was here that Copernicus clearly failed to discover a satisfactory alternative. He provided a neater geometry of the heavens, but it was one which made nonsense of the reasons and explanations that had previously been given to account for the movements in the sky.”\*

PROBLEM 7-3. Read Milton's *Paradise Lost*, Book VIII, lines 1-202, in which Adam and Raphael discuss the two systems from a classic point of view which may well have mirrored the general opinion of educated Englishmen until the full impact of Newton's work was established.

(c) Third, argument arose from the oft-mentioned lack of observable parallax among fixed stars. Copernicus' only possible (and correct) reply was still unacceptable because it involved expanding the celestial sphere to practically an infinite distance away from the earth. This is to us no longer intellectually difficult, but it was a manifestly absurd notion at the time. Among many reasons for this, perhaps the intensely religious mind of the period could not be comfortable with a threatening Hell so close below the ground and a saving Heaven so infinitely far above. If we consider for a moment, it is evident that as a consequence of this attitude an actual observation of annual parallax at that time would not necessarily have settled the dispute in favor of Copernicus, for an obvious and simple Ptolemaic assumption of an additional yearly epicyclic motion superposed on the great daily rotation of the fixed stars would explain away such an observation in full accord with the geocentric system.

(d) A fourth important argument against the Copernican astronomy at the time was that apart from its powerful simplicity it offered to contemporary astronomers no other overruling *scientific* advantages over their geocentric astronomy, i.e., there was then no important observation which was explainable only by one and not by the other, no experiment to pit one against the other in a clear-cut deci-

\* H. Butterfield, *Origins of Modern Science*, p. 27. London: G. Bell & Sons (1950).

sion. Copernicus introduced no fundamentally new experimental facts into his work, nor was the accuracy of his final predictions significantly better than previous estimates.

To us it is clear, although it did not enter the argument then, that the scientific content of both theories, the power of prediction of planetary motion, was about the same at that time. As Francis Bacon wrote in the early 17th century: "Now it is easy to see that both they who think the earth revolves and they who hold the *primum mobile* and the old construction, are about equally and indifferently supported by the phenomena." In our modern terminology we would say (although this is not what Bacon had in mind) that the rival systems differed mainly in the choice of the coordinate system used to describe the observed movements. As measured with respect to the earth, the sun and stars *do* move, and the earth, of course, is at rest. On the other hand, measured with respect to the sun, the earth is not at rest. Any other system of reference, for instance one placing the moon at the origin of the coordinate system and referring all motions to that new center of the universe, is quite conceivable and equally "correct," although, of course, evidently most complex. The choice of a particular frame of reference should, as we now know, depend mainly on which will yield the simplest solution for a given problem. Therefore we cannot speak of such a choice as "true" or "false," but rather as convenient or inconvenient.

Consequently, every newly discovered peculiarity of planetary motion could have been accommodated in some geocentric representation as certainly as in a heliocentric one. In short, neither Aristarchus nor Copernicus could have hoped to convince people to whom the beauty of simplicity and ease of application of a theory were (understandably) less important than their own meaningful and by no means unworkable systems.

**7-5 Historic consequences.** Eventually the vision of Copernicus did triumph; in a moment we shall follow the work leading to the universal acceptance of the heliocentric theory, even though his specific *system* of uniform circular motions will have to be sacrificed along the way. We shall see that the real scientific significance and the reason for the ultimate glorification of Copernicus' work lie in a circumstance he would never have known or understood—in the fact that only a heliocentric formulation opens the way for an integration of the problem of planetary motion with the simple laws of "ordinary" (terrestrial) mechanics as developed during the following 150 years—

a synthesis of two sciences, even of two methods, which was achieved through Newton's Theory of Universal Gravitation. Consequently, it became possible to explain, on the basis of the supposition that the earth turns and travels, such diverse phenomena as the diurnal and the annual apparent motion of the stars, the flattening of the earth, the behavior of cyclones, tradewinds, and gyroscopes, and much else that could not be bound together into so *simple* a scheme in a geocentric physics.

But apart from this historic triumph, the memory of Copernicus is hallowed for two more reasons. First, he was one of those giants of the fifteenth and sixteenth centuries who challenged the contemporary world-picture and thereby gave life to that new and strange idea which later was to grow into science as we now know it. Secondly, his theory proved to be a main force in the gathering intellectual revolution which shook man out of his self-centered preoccupations. Out with the Ptolemaic system has to go also the self-confident certainty that a symbolic uniqueness about our position shows that man is the sum and summit. If this certainty still exists, at least it cannot claim that science backs it up.

An important purpose of our story of planetary motion to this point has been preparation for the following digression before returning to our main topic. We may now ask briefly: By what standards shall we judge a scientific theory?

### Further Reading

A. Armitage, *Sun Stand Thou Still*. New York: Henry Schuman (1947). Also published as *The World of Copernicus*, Mentor Book M65. The life and work of the great astronomer.

H. Boynton, *The Beginnings of Modern Science*. New York: Walter J. Black (1948). Brief biographies and excerpts, e.g., from Copernicus, pp. 5-11.

H. Butterfield, *op. cit.* (Ch. 2), Chapter II.

W. S. Knickerbocker (ed.), *Classics of Modern Science*. New York: Alfred A. Knopf (1927). Biographical notes and long excerpts. Chapter II, excerpt from Copernicus. Also J. W. Knedler, *op. cit.*, pp. 49-72. Also S. Commins and R. N. Linscott, *Man and the Universe: The Philosophers of Science*. New York: Random House (1947), pp. 43-69.

E. Rosen, *Three Copernican Treatises*. New York: Columbia Univ. Press (1939). Includes translation of "Commentariolus," Copernicus' early work.

W. T. Sedgwick and H. W. Tyler, *A Short History of Science*. New York: Macmillan (2d ed. 1939). A useful book. See Appendix C in the 1929 edition for Copernicus' revealing dedication of his main work.

H. Shapley and H. E. Howarth, *A Source Book in Astronomy*. New York: McGraw-Hill (1929). Excerpts, e.g., from Copernicus, pp. 1-12.

## CHAPTER 8

### THE NATURE OF SCIENTIFIC THEORY

As we now turn to the standards by which to judge scientific theories—and the rival planetary theories might serve as models—it is obvious that we cannot pretend to lay down criteria by which a working scientist should check his own progress during the construction of a theory. This cannot be done any more meaningfully than constructing a scientific method for his use. What we shall do here is something else, namely, to attempt a clarification of our thoughts on how we, as students of science and its development, may evaluate a particular theory, past or present.

**8-1 The purpose of theories.** We have argued that the task of science, as that of all thought, is to penetrate beyond the immediate and visible to the unseen, and thereby to place the visible into a new, larger context. For like a distant floating iceberg whose bulk is largely hidden under the sea, only the smallest part of reality impresses itself upon us directly. To help us grasp the whole picture is the supreme function of theory. On a simple level, a theory helps us to interpret the unknown in terms of the known. It is a conceptual scheme which we invent or postulate in order to explain to ourselves, and to others, observed phenomena and the relationships between them, thereby bringing together into one structure the concepts, laws, principles, hypotheses, and observations from often very widely different fields. These functions may equally well be claimed for the hypothesis itself. In truth, we need not lay down a precise dividing line, but might regard theory and hypothesis as differing in degree of generality. Therefore at one extreme we might find the *limited working hypothesis* by which we guide our way through a specific experiment, placing at the other end of the spectrum the *general theory*, which guides the design and interpretation of all experiments in that field of study.

Examples of general theories suggest themselves readily, even if we decide, perhaps rather arbitrarily, to use the word “theory” only for those few historic and general schemes of thought like the theories of planetary motion, of universal gravitation, of the nuclear atom, and the like. Galileo’s theory of projectile motion welded together the laws of uniformly accelerated motion, including free fall and the

principle of superposition of velocities, to produce one conceptual scheme, one over-all method of attacking, predicting, and interpreting every conceivable problem involving bodies moving under the influence of a constant force. Similarly, when Charles Darwin had pondered the results of such fields as paleontology and comparative anatomy, he could find an inspired connection in widely different types of observations, explaining previously separate truths in different disciplines as part of one grand scheme of thought, his theory of evolution.

PROBLEM 8-1. Examine some one theory discussed in other courses with which you are thoroughly familiar—perhaps Gregor Mendel's theory of heredity, the Marxian theory of society, Sigmund Freud's theory of the unconscious, or a theory of "business cycles." Differentiate clearly in each case between the *data* available to the theorizer on one hand and the *hypotheses* used by him to explain the data on the other.

PROBLEM 8-2. Do the same for Ptolemy's and Copernicus' theory of planetary motion.

If we inquire more deeply into the main purposes of theories, we find that there are, on the whole, three functions. Let us summarize them here briefly.

(1) A theory correlates many separate facts in a logical, easily grasped structure of thought. By correlating, by putting in juxtaposition and order previously unrelated observations, we can understand them; we always explain by pointing to a relationship. The Platonists explained the motion of planets by relating planets to certain "necessary" attributes of divine beings. Ptolemy and Copernicus explained it by relating the planetary paths to the mathematical combination of certain more or less simple geometrical figures. And not only will a fruitful theory explain the laws which it relates within its framework, it also will show where and why these laws may not hold precisely in practice, as, for example, Galileo's specific use of his theory of projectile motion to explain why projectiles are not expected actually to follow parabolas for very large paths.

It is a primitive fact about human beings that when they look at the moving parts of some intriguing new toy or gadget, they usually try to visualize a mechanical model of the hidden mechanism, to explain by one theoretical system the separate motions observed. This type of activity seems to be a necessity for the human mind. It is as if the intellect were restless until it can subordinate individual events to more inclusive schemes. It may be a matter of economy of thought, for a good theory allows us to grasp, remember, and

deduce a large number of otherwise elusive facts. (Think of the wealth of facts summarized by the Copernican theory.)

And the simplest theory in physics is often based on a mechanical model. We recall Lord Kelvin's famous statement, "If I can make a mechanical model then I can understand; if I cannot make one, I do not understand." This is, of course, not true for all scientists or all branches of science. Conceptual schemes can certainly not always be phrased in mechanical models, and furthermore there are several famous examples in the history of physics which prove that at times progress may be seriously delayed by a too strong belief in a mechanical model. Particularly, the last hundred years have shown that such models, like all analogies, while often helpful as guides to the imagination, can cause it to fall into dangerous traps. For example, it undoubtedly was—and is—easier to think of light as a vibration moving through a material ether rather than as energy propagated through the void, and yet the ether picture carried with it some eventually inescapable fallacies. Similarly, we owe modern chemistry in large part to the simple, almost naive pictorial representation which John Dalton used to order his thoughts about atoms and molecules but, again, a great deal of early trouble in his atomic theory could be traced to his too concrete and prematurely complete mental picture, as we shall see in the section on the chemical atom.

In fact, it has sometimes been suggested that the science of the ancients, in relying frequently on analogies involving the behavior and drives of living beings, was *organismic*, whereas science from Newton on became *mechanistic*, to turn finally in our century more and more to abstract, *mathematical* sets of mental images. But neither is the mathematical type of model free from the dangers inherent in the mechanistic science. Yet our thoughts proceed, as it were, on crutches, and so depend on these schemes and pictures no matter how incomplete.

(2) Whether general or limited, theories and hypotheses are expected to suggest new relations, to start the imagination along hitherto unsuspected connecting paths between old and new facts. In this sense even the theories that were eventually abandoned or found altogether faulty, such as the Ptolemaic theory of planetary motion, the phlogiston theory of chemistry, or the caloric theory of heat, played a vital and positive role in the life of science, for they tended to propose problems, focus the attention, and integrate the effort of many scientists along the same lines rather than allowing it to be completely dispersed in random directions. Even a fallacious

theory, if actively and widely pursued, can soon lead to the key observations needed for a better theory, for, as Francis Bacon remarked, truth arises more easily from error than from confusion.

PROBLEM 8-3. The following table indicates the period of revolution about the sun (the so-called "sidereal period") of the planets in the Copernican system. Instead of Copernicus' values we shall list approximate modern results.

	<i>Mercury</i>	<i>Venus</i>	<i>Earth</i>	<i>Mars</i>	<i>Jupiter</i>	<i>Saturn</i>
Sidereal period (days)	88	225	$365\frac{1}{4}$	687	4330	10,760

(a) Examine these values and the order of spheres in the two systems (consult the diagrams) and then explain why Copernicus felt that there was exhibited order and harmony in his scheme, "a precise relation between the motion and sizes of the orbs." (b) If a new planet were discovered with a sidereal period of about 1830 days, what would the theory *suggest* about the limits of its orbit? If it should turn out that the actual orbit is larger than Jupiter's, is the whole theory wrong? Explain.

PROBLEM 8-4. Compare the orbit of Venus in the two rival theories by referring to Figs. 6-6 and 7-2. If it were found that Venus was not self-luminous but obtained its light from the sun, what would these two theories suggest concerning the nature of the shadow on Venus that might be seen at different times through a telescope from the earth? (Hint: where would the sun have to be with respect to Venus and the earth when we see Venus fully illuminated? Make sketches of planetary position.)

(3) The foregoing paragraphs hint at a further purpose of theories: the prediction of specific new observable phenomena and the solution of practical problems. In Shakespeare's great tragedy, Hamlet's theory of his uncle's guilt yields him a prediction of how the suspect will react to the little play-within-the-play; conversely, the expected reaction proves Hamlet's theory to be very likely correct, and so gives him the desired solution to the problem of the old king's death. In astronomy and physics, where the problems and predictions needed are usually numerical ones, the theories are, as a rule, quantitative. The problems to be solved may be quite abstract, e.g., why the planets sometimes seem near, at other times far, or may be very practical ones. For example, just as some of the earliest speculations on celestial motions seem to have been prompted by the need to make calendars and to predict eclipses, so one of the problems Copernicus' theory helped to solve was the exact length of the year and of the lunar month, data needed for the revision of the calendar.

PROBLEM 8-5. Read Copernicus' treatise in extensive extract (e.g., Knedler's book) and determine to what extent his work seems to exhibit these three purposes of theory.

**8-2 Criteria for a good theory in physical science.\*** When we examine why scientists as a whole, in the historical development of science, have favored or rejected a given theory, we may discern a few criteria which seemed to have implicitly or unconsciously dominated the slowly evolving process of decision. (But by no means must we suppose that a theory is necessarily ever rejected solely because it fails to answer perfectly to one or another of the questions in the listing that follows.)

Three qualifications have already been cited. (1) A fruitful theory correlates many separate facts, particularly the important prior observations, in a logical, preferably easily grasped structure of thought. (2) In the course of continued use it suggests new relations and stimulates directed research. (3) The theory permits us to deduce predictions that actually check with experience by test, and it is useful for solving long-standing puzzles and quantitative problems.

To illustrate: In Einstein's work on relativity, he first shows that those observations explained by contemporary physics are not in conflict with his brief set of assumptions, but that a large variety of facts from previously separate fields can all be deduced from these assumptions, thereby gathering them into one structure. Secondly, he demonstrates that a few hitherto unlooked-for effects should exist (such as the deflection of light beams passing near large masses) and he commands the experimental astrophysicists to search for these phenomena, for he knows and confesses that the theory depends on this verification. And lastly, not only does he use his theory to explain some old, previously unexplainable observations, e.g., the "erratic" component of the planetary motion of Mercury, but his work has also found amazingly useful applications in almost all branches of physics, from optics to nuclear physics.

The history of science has shown us that prosperous theories frequently have some of the following additional properties:

(4) When the smoke of initial battle has lifted, the successful theory often appears to have simple and few hypotheses—to the extent that the words "simple" and "few" have any meaning at all in science. The test of a theory comes through its use; hence the

---

\* This material is based, in part, on unpublished lecture notes of Professor E. C. Kemble.



survival value of economy. A theory which needs a separate mechanism for each fact that it wishes to explain is nothing but an elaborate and sterile tautology.

(5) Ideally, the assumptions should be plausible—to contemporary scientists, of course—even if they are not immediately subject to test; and the whole tenor of the theory should not be in conflict with current ideas. Where this could not be arranged, the theory faced an often stormy and hostile reception and had to submit to long and careful scrutiny before its general acceptance. Therefore, it did not always grow as quickly and widely as one might have hoped. On the other hand, this cannot be helped, if nature disagrees with current common-sense preconceptions; in fact, precisely these preconceptions may block the road to the core of the difficult problems. As the work of Copernicus and Galileo has already shown us (a strange truth which will become even more evident in subsequent discussions), major advances in scientific theories often depend on the daring and persistence of one investigator who questions the obvious and stubbornly upholds the unbelievable.

This problem of the plausibility of new theories is difficult and fascinating. As Whitehead reminds us, all truly great ideas seemed somewhat absurd when first proposed. Perhaps we call them *great* ideas just because it takes an unusual mind to break through the pattern of contemporary thought and to discern the truth in implausible, “absurd” forms.

Why then do we not drop the tentative requirement of conformity? Why did almost all great innovators in science, from Copernicus to Einstein, initially meet with skepticism—or worse—from most of their colleagues? Is it just unreasoning conservatism on the part of the scientific fraternity? Not at all. In this little collection of cases, we are discussing only revolutionary theories, but we must not allow this to distort our perceptions: great revolutionary ideas arise only very rarely, compared with the large number of workable or even great ideas conceived within the *traditional* setting, so that the individual scientist is naturally predisposed to favor the traditional type of advance which he knows and believes in from personal experience. He, rather like Galileo’s opponents, quite rightly must defend his fundamental concept of nature against large-scale destruction, particularly at the earlier stages, where the innovators cannot present very many results and confirmations of their new ideas. Indeed, sometimes the discoverers themselves are so strongly committed to the established ideas, so startled by the implications of their own work, that they predict the storm of condemnation or even simply

fail to draw the final conclusions. A beautiful and now famous instance of this sort appears in the original publication (1939) by Hahn and Strassmann, who are usually credited with the experimental discovery of nuclear fission. At that time it was still axiomatic in the thinking of all scientists that atomic nuclei are indivisible, that bombardment with small particles (e.g., neutrons) may at best change their internal atomic structure only superficially or slightly. But these two men, after firing neutrons into pure uranium, were left with material which by chemical test was proved to contain barium (Ba) and other elements whose atoms are about half as large as uranium atoms—"evidently" as we would say now, the result of splitting uranium atoms. But in an almost agonized expression of the labor pains attending the birth of all great recognitions, Hahn and Strassman could not quite dare to accept the evidence of their own chemical experiments, and so they wrote: "As 'nuclear chemists,' in many ways closely associated with physics, we cannot yet bring ourselves to make this leap in contradiction to all previous lessons of nuclear physics. Perhaps, after all, our results have been rendered deceptive by some chain of strange accidents."\*

Later, after having developed our view that science has its laws of evolution analogous to those governing the evolution of species, we shall appreciate more fully the significance of the bitter and long struggles that new theories may create within science. *The fitness of truths is most advantageously shaped and most convincingly demonstrated in vigorous contest.* The situation is, after all, not so very different in other fields. The predominant religious and social concepts of our time have not all developed quietly or been accepted spontaneously. If the physical sciences sometimes seem to have progressed so very much faster, if the recent struggles have been relatively short, we might, to a great extent, credit this to the emergence, from the 17th century onward, of a more or less tacit agreement among scientists by what standards to judge a new conceptual scheme—above all, of course, the pragmatic test of predictions.

Even so, scientists are, above all, human; a really new and start-

\* Contrast this caution and the willingness to be proved wrong by their fellow scientists, expressed in the previous quotation, with the sensational claims and impatient pomp of our modern pseudoscientific best-sellers, one of whose authors begins his treatise with the sentence that the work "is a milestone for Man comparable to his discovery of fire and superior to his invention of the wheel and arch . . . The hidden source of all psychosomatic ills and human aberration has been discovered, and skills have been developed for their invariable cure."

ling idea may in retrospect clearly have fulfilled all these requirements without being widely accepted. Max Planck, with perhaps only a little too much bitterness about his own early struggles, says, "An important scientific innovation rarely makes its way by gradually winning over and converting its opponents: it rarely happens that Saul becomes Paul. What does happen is that its opponents gradually die out, and that the growing generation is familiarized with the ideas from the beginning."

(6) History points out another feature of successful theory: it is flexible enough to grow, and to undergo minor modifications where necessary. But if, after a full life, it eventually dies, it dies gracefully, leaving a minimum of wreckage—and a descendant.

Let us consider these interesting points. In any field a prediction is no more than a guess unless it follows from some theory. And just as the meaning of a prediction depends on a theory, so does the validity of a theory depend on the correctness of the predictions. Like the very laws it correlates, a theory must be based on a finite, though perhaps small, number of observations; yet to be useful it may have to predict correctly a very large number of future observations. Think again of the Ptolemaic theory; it was built on many separate observations of planetary positions, but it yielded enough reasonably correct predictions of positions to be still useful more than 1400 years later. Yet we see at once that physical theories are not likely to last forever. Sooner or later there will appear observations to contradict the predictions, probably in a region for which the conditions had to be extrapolated from the original range of the theory (e.g., the region of very small or very large quantities, or of much more precise measurements than served initially as a basis for the theory), or perhaps as a result of a general widening of the scientific horizon. With ingenuity one can usually find that the old theory can be modified to apply also to the new phenomena. Like an apple tree in an orchard, we keep a theory for the sake of its fruits; when the apple crop becomes poor, we then shall try to save the tree by judicious pruning and the like. In the endangered theory, perhaps a re-examination of some assumptions or concepts will save the scheme. But otherwise the only alternatives are to retain the old theory "as is" with the clear understanding of the restricted range to which it still applies, as before, and to develop a new structure to take care of the new range; or, if the flaws appeared within the proper jurisdiction of the old theory and the assumptions stand challenged beyond repair, to abandon it as soon as a new, better one can be contrived.

These three choices—expansion, restriction, or death—appear in the history of theories and of laws generally. We tend to remember and to honor most those schemes which indicated the solution in a crisis, or left some hint of a new start at their demise. In the words of Niels Bohr, “The utmost any theory can do [is] to be instrumental in suggesting and guiding new developments beyond its original scope.”

But finally, lest our brief paragraphs on these criteria of good theories be mistaken for scientific dogma, we should echo the very humble opinion of Einstein on this same topic. He distinguishes between two main criteria, (a) the *external confirmation* of a theory, which informs us in experimental checks of the correctness of the theory, and (b) the *inner perfection* of a theory which judges its “logical simplicity” or “naturalness.” He then qualifies these remarks as follows: “The meager precision of the assertions [(a) and (b) above] . . . I shall not attempt to excuse by lack of sufficient printing space at my disposal, but confess herewith that I am not, without more ado, and perhaps not at all, capable to replace these hints by more precise definitions.”

### Additional Problems

**PROBLEM 8-6.** Investigate the death of a theory from another field with which you are very well acquainted (such as the phlogiston theory in chemistry or the theory of spontaneous generation in biology) in terms of our six criteria. What were the predictions which failed to come true, or the phenomena which contradicted the assumptions of the theory? Could it have been modified to serve for a time? Did a new theory rise directly from the old?

**PROBLEM 8-7.** P. W. Bridgman writes in *Reflections of a Physicist* (Philosophical Library, New York, 1950, p. 252) that in the current flux in ideologies, moral ideas, and other social concepts the intelligent scientist sees “. . . an exemplification of what his physical experience has taught him; namely, ideas are to be inspected and re-examined when the domain of application is extended beyond that in which they arose. In one very important respect he recognizes that the present epoch differs from former epochs in that the enormous increase in invention, bringing peoples nearer together and increasing their command over forces more advantageous to man, effectively provides just that extension in the domain of application of social concepts which he is prepared to expect would demand fundamental revision.” Write a short essay on your opinion about this quotation, and also examine what other guiding ideas you think may be transferred to the field of social studies from the study of physical theories.

PROBLEM 8-8. Was the geocentric theory as set forth by Ptolemy *false*? Was it *true*?

PROBLEM 8-9. Discuss the relative advantages of the Ptolemaic and Copernican theories of celestial motion, comparing both theories according to each of the six criteria developed in the preceding section.

PROBLEM 8-10. Do the same for the Aristotelian and Galilean theories of freely falling bodies (cf. Chapter 2).

### Further Reading

*Note:* For further reading, consult the bibliography at the end of Chapter 14.

## CHAPTER 9

### KEPLER'S LAWS

The theory of planetary motion now develops with gathering momentum. We have reached the years around 1600 A.D. The Renaissance and Reformation are passing. Copernicus has been read by few and accepted by fewer astronomers. Through this silence we hear one striking voice sounding the first cries of a coming battle. The antiorthodox pantheist and evangelizing Copernican, Giordano Bruno, is traveling through Europe, announcing that the boundaries of the universe are infinitely far away and that our solar system is one of infinitely many. For his outspoken heresies he was tried by the Inquisition and was burned at the stake in 1600.

But the seeds of a new science are sprouting vigorously here and there. In England there is Francis Bacon (1561-1626) and William Gilbert (1540-1603); in Italy, Galileo (1564-1642). And at Copenhagen, Tycho Brahe (1546-1601), since the Greeks the first man to bring real improvements into astrometry, spends nearly a lifetime in patient observation and recording of planetary motion with unheard-of precision. His data are often accurate to less than half a minute of arc, more than twenty times better than those of Copernicus, even though the telescope has not yet been invented.

After Tycho's death, his German assistant Johannes Kepler (1571-1630) (see Fig. 9-1) continued the observations and above all the reductions of the voluminous data. Whereas Tycho had developed a geocentric system of his own, which became widely accepted for a time, Kepler was a Copernican. The announced purpose of his work was the construction of better astronomical tables of planetary motion than were then available on the basis of the more dubious data of Copernicus' own period. But Kepler's motivation and main pre-occupation was the perfection of the heliocentric theory, whose harmony and simplicity he contemplated "with incredible and ravishing delight." Particularly at the outset of his long labors he was strongly influenced by Pythagorean and neo-Platonic metaphysics. To him even more than to Copernicus, the clue to God's mind was geometric order and numerical relation, expressed in the features of the simple heliocentric scheme. Among his earliest works we find, typically, an enthusiastic attempt to link the six known planets and their respective distances from the sun with the relationships between the five



FIG. 9-1. Johannes Kepler (1571-1630).

regular solids of geometry. The best one can say for this work was that it served to bring Kepler to the attention of Tycho and Galileo.

In attempting to fit the accurate new data on Mars' orbit to a Copernican system of simple uniform circular motion (*even* if equants were used), Kepler made a terrible discovery—he could not do it! After four years of calculations he still found that the new data placed the orbit just eight minutes of arc outside the scheme of Copernican devices. This is many minutes less than the error which Copernicus knew his own data to contain; therefore Copernicus would not have noticed this deficiency. But Kepler could trust Tycho's unflinching eye and superb instruments, and with an integrity that has become the characteristic attitude of scientists in the face of quantitative fact, he would not allow himself to hide this fatal difference behind some convenient assumptions. To Kepler, these eight minutes meant simply that the Copernican scheme of concentric spheres, epicycles, and eccentrics failed to explain the actual motion of Mars when the observations were made accurately enough.

One might feel tempted to return to the Ptolemaic scheme after all, and in fact it should be able to explain the motion of Mars, but only if we heap device on device. Thus in a sense both theories fail us; one is not quite accurate enough, and the other breaks under its own weight.

**9-1 Kepler's first law.** Kepler must have been stunned, for after all he was a convinced Copernican. After years of continuous labor he at length discovered how the Copernican theory could be amended to make it applicable to both the old and the new observations—by dropping the one assumption which most bound it explicitly to the doctrines derived from ancient Greece. When he plotted the paths of planets on the basis of a heliocentric representation, and when instead of trying to interpret these paths as combinations of different motions he considered the shape of the orbits themselves, he found that they corresponded to a simple type of figure whose properties had been known to mathematicians since the second century B.C.—the ellipse. If, therefore, the ellipse is recognized as the natural path of celestial bodies, then we have a geometrically truly simple world-scheme. "Planets move in elliptical paths, with the sun at one focus of the ellipse." This may be our statement of Kepler's first of three great laws, the *Law of Elliptical Paths*. Note the similarity between this and Galileo's work on trajectories. We recall that Galileo explained how projectiles moved by pointing out that they all followed paths known to the mathematicians as parabolas. Kepler again matches a physical path against a mathematical curve, and so simultaneously describes and begins to explain planetary motion.

We have here an example of the enormous change in outlook in Europe that had begun more than two centuries before. More and more, events ceased to be regarded as *symbols* and were allowed to stand for themselves. Man ceased to be preoccupied with anthropomorphic riddles in an organismic world, and slowly became a factual observer and theorizer in a mechanistic world. Without this new attitude, there could have been no modern science, for if we are to start our science from experimental observables, we must have faith that we are dealing with the raw material of experience, not with symbols of complex mysteries. We must become enthusiastic about the observable world for its own sake, we must attain a tacit faith in the meaningfulness of Nature and her direct accessibility to our understanding, before we can expect generations of scientists to devote themselves to the minute and often most tedious quantitative



investigations of nature. In this sense Kepler's work heralds the change toward the modern scientific attitude—to regard a wide variety of phenomena as explained when they can all be described by one simple, preferably mathematical, pattern of behavior.

It seems at first astonishing that Kepler should be the one to follow this new road. He had begun his career as a symbol-seeking mystic, but now we may discern a reflection of the great change in his complex soul: he fashions his physical laws—and looks for his symbolism afterwards. Philosophical speculation, still rather colorful, follows factual analysis, not the reverse; and to this day many scientists have found it quite possible to reconcile their physics and their personal philosophy on the basis of this sequence.

Kepler's first law, by amending the Copernican heliocentric theory, gives us a wonderfully simple mental picture of the solar system. Gone are all the epicycles and eccentrics; the orbits are clean and elliptical. Figure 9-2 shows a schematic representation of the present conception of the solar system—in essence Kepler's, but with the addition of the planets Uranus, Neptune, and Pluto, discovered much later. An attempt has been made to represent on the figure the relative sizes of the planets; the approximate relative sizes of the orbits, to a different scale, are indicated also, but since most of these ellipses are quite nearly circular, they are shown as circles where possible. All orbits are nearly in the same plane, except for the more pronounced tilt of Pluto's (which means that Neptune's and Pluto's paths do not cross at any point in space, although Fig. 9-2 may give that impression).\*

PROBLEM 9-1. The important properties of the ellipse. The equation for an ellipse in rectangular coordinates is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

where  $a$  and  $b$  are respectively half the major and half the minor axis. On graph paper, construct a large ellipse by deciding on some constant values for  $a$  and  $b$  and by then solving for a few  $\pm y$  values for assumed  $\pm x$  values. (Note that if  $a = b$ , the figure is a circle.) The distance  $c$  from the origin of the coordinates to one or the other focus ( $F_1$  or  $F_2$ ) is given by

$$c = \sqrt{a^2 - b^2}.$$

---

\* See F. L. Whipple's *Earth, Moon and Planets* [Philadelphia: Blakiston Co. (1941)], for a readable account of the structure of the solar system. A brief summary is in Shapley, Wright, and Rapport, *Readings in the Physical Sciences*, pp. 103-108 [New York: Appleton-Century-Crofts (1948)].

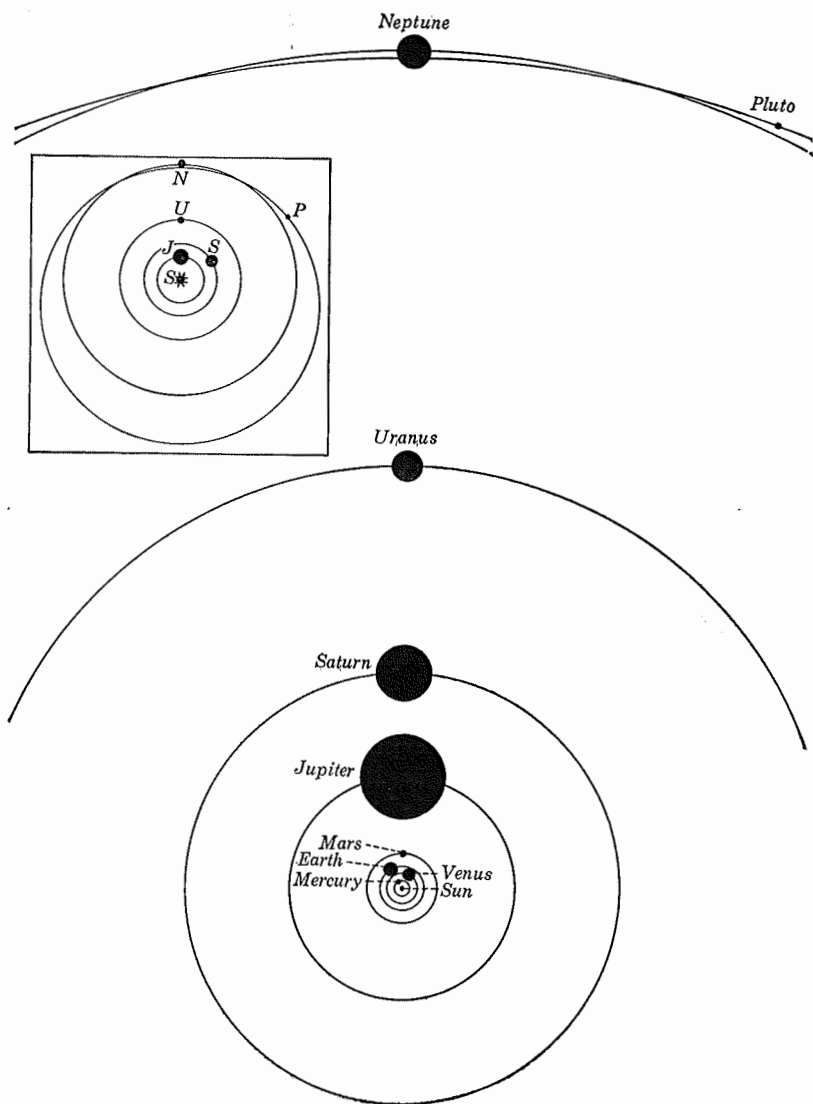


FIG. 9-2. Schematic outline of solar system, showing relative sizes of the orbits and, to a different scale, the relative sizes of the planets. The insert traces Pluto's full orbit.

Enter the foci on your ellipse. The eccentricity\*  $e$  is a quantity indicating by how much an ellipse differs from a circle. By definition,

$$e = \sqrt{1 - \frac{b^2}{a^2}}$$

Find  $e$  for your ellipse. Test by measurements that for a point  $P$  anywhere on the ellipse,  $PF_1 + PF_2 = a$  constant. Lastly, note that for a given set of coordinates an ellipse in a given plane is completely defined (i.e., can be correctly drawn) if just its values for  $a$  and  $b$  are given; these two data suffice. (Fig. 9-3.)

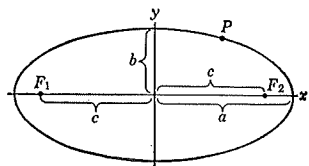


FIG. 9-3. Ellipse, with foci at  $F_1$  and  $F_2$ .

PROBLEM 9-2. Determine the approximate value of  $e$  for Pluto from measurements on Fig. 9-2. (Note:  $e$  for Mercury is about 0.2, and for the remaining planets even less.)

After Kepler's initial joy over the discovery of the law of elliptical paths had subsided, he may have asked himself the obvious question, "Is it not rather mysterious that of all possible types of paths the planets all select just the ellipse?" We could understand Plato's predisposition for uniform circular motions, but we cannot readily comprehend nature's insistence on the ellipse! It is no longer a man-made mystery, but therefore all the darker. In fact, there was no rational answer until a reticent genius in England almost eighty years later showed the law of the ellipse to be one of the many surprising consequences of a much more far-reaching law of nature. We shall follow his reasoning in a few moments.

If for the present we accept Kepler's first law purely as a summary of observed facts—an "empirical" law—we note that by describing the paths as elliptical, the law gives us all the possible positions of a given planet—but it does not tell us *when* it will be at any of these positions; it talks about the shape of the orbit, not the changing speed along the orbit. This makes the law rather useless for an astronomer who wishes to know where he should expect to look for a planet at a given time. For an anxious moment it seems that Kepler has sacrificed the possibility of astronomical prediction for the sake of the simplicity of path. Less than nothing is gained, because the other theories gave at least the possibility of fairly good prediction of a planet's possible positions *and* of its time of appearance at each possible position.

\* Do not confuse with the word "eccentric" as used in Section 6-4.

**9-2 Kepler's second law.** Kepler was perfectly aware that his first law alone was not enough to justify a jubilant overthrow of the apparatus of epicycles and eccentrics. He set out to discover a law which would relate the speed of a planet in any position of its orbit with the speed in any other position. If such a relation could be found, the motion of any one planet would be specifiable by just a few separate figures: two data to specify the ellipse (e.g., the major and minor axes); one more to give the speed at a certain portion of the path, for instance when the planet is nearest the sun, "at the perihelion"; and a few additional data to place this ellipse at the correct tilt with respect to the others. Thus we could hope to summarize all features of planetary motion in a compact, elegant way.

But does the hoped-for relation between speed and path actually exist? There appears so far no necessary reason that it should. Kepler is said to have been in ecstasy when by sheer labor and ingenuity he finally read the needed second law out of his own voluminous data. Well he might have been; his whole labor would have been of little use without this additional discovery.

Figure 9-4 indicates what he noticed. We are looking down on a greatly exaggerated representation of a planetary path around the sun  $S$  at one focus of the ellipse. The planet may be our earth; it is found to move from position  $P_1$  to  $P_2$  (say during a week in our winter). During an equal time interval in the spring the journey proceeds from  $P_3$  to  $P_4$ , and from  $P_5$  to  $P_6$  in summer. The implication from the drawing is that the planet's speed along the orbit is greatest when nearest the sun, less when more distant. For the earth, the maximum orbital speed is 18.8 miles per second at perihelion, the minimum 18.2 miles per second at aphelion.

This gratifying regularity turns into a law by the deeper recognition that by measurement the area of the figure  $SP_1P_2$  is the same

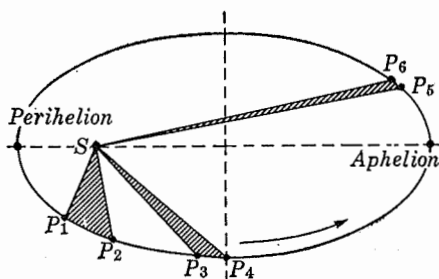


FIG. 9-4. Elliptical path of planets around sun  $S$  (at left focus), illustrating Kepler's Second Law (eccentricity much exaggerated).

as for  $SP_3P_4$  and for  $SP_5P_6$ , etc. During a given time interval a line from the planet to the sun sweeps out an equal area anywhere along its elliptical path. This second law, or *Law of Equal Areas* applies to all planets.\*

**9-3 The meaning of the second law.** Before we use this law to give us directly the needed relation between path and speeds, we must look at some mathematical consequences. Assume that the time taken from  $P_1$  to  $P_2$ , or  $P_3$  to  $P_4$ , etc. is always the same, and call it  $t$ . Let it be short, so that the distance from  $S$  and the speed at  $P_1$  differ very little from the values at  $P_2$ , at  $P_3$  very little from those at  $P_4$ , etc. Then each of the three areas in Fig. 9-4 is approximately a circular sector. Plane trigonometry gives the area of such a figure as  $\frac{1}{2} \times \text{radius} \times \text{arc}$ . Let us say

$$\overline{SP}_1 \doteq \overline{SP}_2 \equiv r_1; \quad \overline{SP}_3 \doteq \overline{SP}_4 \equiv r_3; \quad \overline{SP}_5 \doteq \overline{SP}_6 \equiv r_5.$$

Then Kepler's second law states:

$$\frac{1}{2} r_1(\widehat{P_1P_2}) = \frac{1}{2} r_3(\widehat{P_3P_4}) = \frac{1}{2} r_5(\widehat{P_5P_6}). \quad (9-1)$$

We can convert this set of equations into another form and so get a quite new meaning for this law. It is certainly permissible to multiply *each* expression in Eq. (9-1) by the factor  $2/t$ ; thus

$$r_1 \frac{\widehat{P_1P_2}}{t} = r_3 \frac{\widehat{P_3P_4}}{t} = r_5 \frac{\widehat{P_5P_6}}{t}. \quad (9-2)$$

However,  $\widehat{P_1P_2}/t$  is the average speed  $\bar{v}_1$  of the planet between  $P_1$  and  $P_2$ , etc. If the interval  $t$  is chosen suitably small,  $\bar{v}_1$  approaches sufficiently close to the instantaneous speed  $v_1$  of the planet at that portion of the path, and we may write

$$r_1 v_1 = r_3 v_3 = r_5 v_5. \quad (9-3)$$

Even though this equation becomes rather more complex if we desire accuracy and consider the actual elliptical path instead of the approximation used here, we may nevertheless deduce a very important result from Eq. (9-3):

$$v_3 = v_1 \frac{r_1}{r_3}; \quad v_5 = v_1 \frac{r_1}{r_5}, \quad (9-4)$$

---

\* Also to the moon in its path around the earth insofar as this is elliptical, and to all other satellites about their own planets. A restriction on our use of the second law is that, in order to facilitate the mathematical part of the discussion, we shall apply this law to the simple case of only those ellipses whose shapes are not too far from the circular, a good approximation to most actual orbits.

or, in short, if we know by observation of the planet its speed  $v_1$  at portion  $P_1P_2$  or distance  $r_1$ , we can then predict the speed  $v_x$  at any other distance  $r_x$ . And knowing now how to calculate the speed of the planet at each point on its ellipse, we can predict either how long it will take to reach a given spot or how far it will have gone in a given time interval. Thus Kepler's second law re-establishes the possibility of astronomical prediction, and in a simpler manner than through the multiplicity of geometric devices of the previous theories of planetary motion.

**9-4 Kepler's third law.** Kepler's first and second laws were published together in 1609 in his *New Astronomy*. But he was still unsatisfied with one aspect of his achievement: it had not revealed any connection between the motions of the different planets. So far, each planet seemed to have its own elliptical orbit and speeds, but there appeared to be no over-all pattern for all planets. Nor was there any good reason why one should expect such a relationship. Kepler, however, was convinced that on investigating different possibilities he might hit on a simple rule linking the whole solar system.

This conviction, so strong that it seems to us like an obsession, was partly a remnant of his earlier numerological preoccupations. But it also indicates a deep undercurrent running through the whole history of science: the belief in the simplicity and uniformity of nature. It has always been a source of inspiration, helping scientists over the inevitable unforeseen obstacles, and sustaining their spirit during periods of long and fruitless labor. For Kepler it made bearable a life of heart-breaking personal misfortunes,\* so that he could write triumphantly, on finally finding the great third law:

“. . . after I had by unceasing toil through a long period of time, using the observations of Brahe, discovered the true distances of the orbits, at last, at last, the true relation . . . overcame by storm the shadows of my mind, with such fullness of agreement between my seventeen years' labor on the observations of Brahe and this present study of mine that I at first believed that I was dreaming . . .”

The law itself, in modern terminology, states that if  $T$  be the sidereal period of any chosen planet (i.e., the time for one complete

\* Movingly described in Sir Oliver Lodge's *Pioneers of Science*. London: Macmillan (1893).

orbital revolution about the sun), and  $\bar{R}$  be the mean radius of the orbit of that planet,\* then

$$T^2 = K(\bar{R})^3,$$

where  $K$  is a constant having the same value for *all* planets. But if  $T^2/(\bar{R})^3$  is the same for all planets, we can calculate its numerical value for *one* planet (for the earth:  $T_E = 1$  year,  $\bar{R}_E \doteq 9.3 \times 10^7$  miles), and thereafter we shall always be able to compute for any other planet its  $T$  if its  $\bar{R}$  is given, and vice versa.

PROBLEM 9-3. Using the third law and the data for  $T_E$  and  $\bar{R}_E$  for the earth, compute values of  $\bar{R}$  for a few other planets, taking the corresponding values of  $T$  from the quotation from Copernicus' *Revolutions* in Section 7-2.

Appropriately, Kepler's third law is often called the *Harmonic Law* because it establishes a beautifully simple relationship among the planets. From this peak we may survey our progress so far. Starting from the disconnected multitude of Ptolemaic gears we have reached a heliocentric formulation that views the solar system as a simple, logically connected unit. Our mind's eye grasps the Keplerian universe at one glance, and recognizes the main motions as the expression of simple mathematical laws.

**9-5 The new concept of physical law.** With Tycho's work, his own observations, and his powerful three laws, Kepler was able to construct the long-needed accurate tables of planetary motion which remained useful for a century. We honor him for all these great accomplishments in astronomy (and these are only a few of the achievements of this prodigious man) but we must also not forget to single out two features of his work which had a profound effect on the development of all physical sciences. One, which has been discussed, is the rather new attitude to *observed facts*. We noted the change in Kepler's work from his earlier insistence on a geometric model and its form as the main tool of explanation, to the discussion of the movement itself and of the numerical relations underlying it. The other is his successful attempt to formulate physical laws in mathematical form, in the language of geometry and algebra. From

---

\* The value of  $\bar{R}$  for an elliptical path is half the distance of a straight line from perihelion to aphelion (see Fig. 9-4); conveniently, most planetary paths are nearly enough circular so that  $\bar{R}$  is then simply the radius of the circular orbit.

then on, from the period of Kepler and Galileo, the *equation* develops naturally as the prototype of the laws in physical science.

In this sense, Kepler's science was truly modern; he more than anyone before him bowed to the relentless and supreme arbiter of physical theory, namely, the evidence of *precise* and *quantitative* observation. But in the Keplerian system the planets no longer are considered to move in their orbits because of their divine nature or the continual effort of angels, as in scholastic teaching, nor because their spherical shapes themselves were self-evident explanation for their circular motion, as in Copernicus' thoughts, and so we are left without any physical agency to "explain" or give plausibility to the planetary motion so well described in these three laws. Kepler himself felt this need of backing up mathematical descriptions with physical mechanism. While it was left to Newton to assign this task to gravitational forces, and thereby to tie Kepler's three laws together with the heliocentric conception and the laws of terrestrial mechanics in a monumental synthesis, Kepler entertained a then quite promising hypothesis: the recent work of Gilbert on magnetism had greatly intrigued him, and his rich imagination could picture magnetic forces emanating from the sun to "drive" the planets in their orbits. (Newton later felt it necessary to prove in some detail that this hypothetical agency would not account for the quantitative observations.)

A final word about Kepler's place in present-day astronomy: Three centuries of telescopic observations have shown that Kepler's laws hold within well-understood and usually small deviations. The laws have also been found to have a much broader scope; for example, in atomic physics we shall use them to help us describe the paths of electrons around nuclei. Fruitful ideas, it is evident, are bound neither by time nor by space, nor yet by the artificial boundaries of academic fields.

### Additional Problems

PROBLEM 9-4. Evaluate Kepler's contribution in terms of our criteria for a good theory as far as you can.

PROBLEM 9-5. Trace the changes and evolution of the Copernican heliocentric theory in the hands of Kepler, and then compare this with the development and expansion of some important theory from another field familiar to you (e.g., the Theory of Evolution, from its exposition in 1859 to the recent modifications of evolutionary genetics).

PROBLEM 9-6. Does it seem possible that a revision of the geocentric theory can be made, doing away with equants, eccentric motions, etc., and



instead describing the planetary paths *about the earth* directly by simple geometric figures? Is it likely that laws could be found for this simplified geocentric system which relate the paths, speeds, and periods of revolution of the different planets?

### Further Reading

R. H. Baker, *Astronomy*. New York: D. Van Nostrand Co. (1930 and later editions). Excellent introduction.

W. W. Bryant, *Kepler*. New York: Macmillan (1920).

J. L. E. Dreyer, *History of the Planetary Systems from Thales to Kepler*. Cambridge: Cambridge University Press (1906).

W. S. Knickerbocker, *op. cit.* (Ch. 7). Chapter III, excerpt from Kepler.

P. Lenard, *Great Men of Science*. New York: Macmillan (1934). Short biographies, *e.g.*, of Copernicus, Tycho Brahe, Galileo, Kepler.

H. N. Russell *et al.*, *Astronomy*. Boston: Ginn and Co. (Volume I, 1926 and later editions).

H. Shapley and H. E. Howarth, *op. cit.* (Ch. 7). Excerpt from Tycho Brahe (pp. 13-19) and Kepler (pp. 29-40). Also H. Boynton, *op. cit.*, pp. 12-26.

## CHAPTER 10

### GALILEO'S CONTRIBUTION TO ASTRONOMY

"There are more things in heaven and earth, Horatio,  
Than are dreamt of in your philosophy . . ."

*Hamlet*, Act I

One of the few friends and fellow-scientists with whom Kepler corresponded and exchanged news of the latest findings was Galileo Galilei. Although the Italian's *scientific* contribution to planetary theory is not so well developed as that of his friend across the Alps, he nevertheless was projected into history as a key figure in this subject. In a sense, Kepler and Galileo complemented each other in preparing the world for the eventual acceptance of the heliocentric theory—the one laying the scientific foundation with his astronomical work, the other fighting the dogmatic objections and, in his work on mechanics, helping to overturn the whole structure of scholastic physics with which the old cosmology was entwined.

For it was Galileo more than any other man who, as we have already seen, challenged the fruitfulness of the ancient interpretation of experience, and focused the attention of physical science on the productive concepts—time and distance, velocity and acceleration, force and matter—and not on qualities or essences, ultimate causes or harmonies, which were still the motivation of a Copernicus and at times the ecstasy of a Kepler. Galileo's insistence, so clearly expressed in his work on freely falling bodies, on fitting the concepts and conclusions to observable facts, on expressing his results in the concise language of mathematics, are now accepted as central achievements, re-enforcing the same traits then emerging from the work of Kepler. But perhaps the greatest difference between the work of Galileo and that of his scholastic contemporaries was his orientation, his viewpoint, the kind of question he considered important. To most of his opponents, Galileo's specific problems were not general enough, since he excluded the orthodox philosophical problems. Then, too, his procedure for discovering truth seemed to them fantastic, his conclusions preposterous, haughty, often impious. There exists an almost perfectly parallel situation between these objections to Galileo's point of view and the outraged derision and even violence initially heaped on the discoverers of new ways of viewing the world in art, e.g., Manet and his impressionists.

**10-1 The telescopic evidences.** Like Kepler, Galileo was a Copernican in a Ptolemaic world. His specific contributions to the development of heliocentric theory are two: the invention in 1609 of a telescope for astronomical work, and the observations he made with its help. The same set of doctrines which then still generally upheld the geocentric system as the only possible choice also required that the celestial objects be "perfect" (spherical and unblemished), yet he could now plainly see spots on the sun and mountains on the moon. Venus, sometimes fully illuminated by the sun and at other times not at all, was found to have moonlike phases. Saturn seemed to carry bulges around its equator. Most surprisingly, Jupiter showed him four (of its twelve now known) satellites or subplanets, adding in a catastrophic manner to the seven scholastically acceptable members of the planetary system, and offering to plain view a miniature solar system with an "obvious" center of rotation far from the earth. The Milky Way resolved into aggregates of individual stars.\* The stars themselves, now visible in much greater numbers, appeared still as pinpoints of light, thus aiding the Copernican argument that they are extremely distant.

However, we must understand that these suggestive observations were to Galileo illustrations of a truth that seemed to him compelling even from a nonexperimental point of view. In his great work *The Dialogue on the Two Chief Systems of the World*, he stresses the one argument as much as the other. Observations alone, says Galileo, do not decide uniquely between a heliocentric and a geocentric hypothesis, "for the same phenomena would result from either hypothesis." This would be close to our modern view, but where we would make a choice between the two mainly on the basis of convenience in use, without regarding one as true and the other as false, Galileo understandably thinks of the earth's motion as "real," real just *because* it seems more reasonable and simplifies the picture. He then adduces several arguments, largely following Copernicus:

"When we merely consider the immensity of the starry sphere in comparison with the smallness of the terrestrial ball . . . and then think of the rapidity of the motion which completes a whole rotation in one day and night, I cannot persuade myself how any-

---

\* Democritus (b. 460 B.C.), one of the originators of the atomistic view of matter, appears to have speculated quite successfully on the nature of the Milky Way and the appearance of the moon. See C. Bailey, *The Greek Atomists and Epicurus*, p. 151, Oxford University Press (1928).

one can hold it to be more reasonable and credible that it is the heavenly sphere which rotates, while the earth stands still." \*

As a second point Galileo reminds his readers that in the geocentric model it is necessary to ascribe to the planets a motion opposite to that of the celestial sphere (why?), again an unreasonable, we might almost say unharmonious or unesthetic assumption. Third, Jupiter's four moons had shown him that there, too, existed the rule that the larger the orbit of the rotating body as reckoned from the center of rotation, the longer the period of rotation (qualitatively, Kepler's third law). This Copernicus had pointed out long before for the case of the planets themselves, and it held true even in the Ptolemaic system—but with this disharmony: in the Ptolemaic system the characteristic periods of revolution around the earth increase from the short one ( $27\frac{1}{3}$  days) for the moon, to the very large one (30 years) for Saturn—and then suddenly drop back to 24 hours for the celestial sphere. In the Copernican system, however, "the rapidity of the periods is very well preserved; from the slowest sphere of Saturn we come to the wholly motionless fixed stars. We also escape thereby a fourth difficulty . . . I mean the great unevenness in the movement of these very stars [on the Ptolemaic hypothesis], some of which would have to revolve with extraordinary rapidity in immense circles, while others moved very slowly in small circles, since some of them are at a greater, others at a less distance from the pole." Fifth, owing to the slightly changing tilt of the earth's axis, the apparent paths of the stars on the celestial sphere change slowly over the centuries, again an improbable or at any rate an unreasonable feature of a geocentric theory that claimed to be based on the immutable, ideal, eternal characteristics of the heavenly bodies. Next, Galileo found it impossible to conceive in what manner the stars could be fixed in the celestial sphere to allow them rapid rotation and even motion as a group (*cf.* last point), while also having them preserve their relative distances so perfectly. "It seems to me much easier and more convenient to make them motionless instead of moving" and to ascribe the apparent motions to that of one "small insignificant body in comparison with the whole universe," the earth itself. Lastly Galileo points out "I could also not explain why the earth, a freely poised body, balancing itself about its center [on the geocentric scheme] and surrounded on all sides by a fluid medium [postulated at the time to provide means for communicating the "force and

---

\* From the translation by Dannemann.

power" needed to keep up the celestial motions] should not be affected by the universal rotation."

In the text, Galileo's hypothetical opponent answers to these points:

"You support your arguments throughout, it seems to me, on the greater ease and simplicity with which the said effects are produced. You mean that as a cause [of the observed motion of planets and stars] the motion of the earth alone is just as satisfactory as the motion of all the rest of the universe with the exception of the earth; you hold the actual event to be much easier in the former case than in the latter."

These are meant to be accusing or scoffing remarks, as though, in modern terminology, one would say, "You are just a relativist, willing to throw to the winds, for the sake of convenience, all the sacred implications of our geocentric hypothesis." And, in fact, the fictional opponent then dismisses all of Galileo's arguments with this quite characteristic counter-argument:

"For the ruler of the Universe, however, whose might is infinite, it is no less easy to move the universe than the earth or a straw balm."

Analyze carefully this part of Galileo's modern-sounding and conclusive reply (and incidentally notice the sly final phrases):

"What I said . . . does not refer to Him who causes the motion, but to that which is moved. . . . If we consider the moving bodies, we must unquestionably regard the motion of the earth as a *much simpler process* than that of the universe. If, furthermore, we direct our attention to *so many other simplifications* which may be reached *only* by this heliocentric theory, the daily movement of the earth must appear much more probable than the motion of the universe without the earth; for according to Aristotle's just axiom, 'It is vain to expend many means where a few are sufficient.' "

PROBLEM 10-1. Examine each of Galileo's telescopic observations separately to determine whether it aided or disproved (a) the geocentric theory as an astronomical system, (b) the heliocentric theory as an astronomical system, (c) the philosophic doctrines behind the geocentric theory. In particular, note Galileo's observations for Venus, and refer to your answers to Problem 8-4. It should be clear that the Ptolemaic system, without some rearrangement, cannot account for all of Venus' moonlike phases. Propose

a modification of the Ptolemaic system which would remove the difficulty while retaining the *geocentric* feature.

In his characteristic enthusiasm, Galileo thought that through his telescopic discoveries everyone would see, *as with his own eyes*, the absurdity of the assumptions which prevented a general acceptance of the Copernican system. But man can believe only what he is ready to believe. In their fight against the new Copernicans the scholastics were convinced that they were surely "sticking to facts," that the heliocentric theory was obviously false and in contradiction with both sense-observation and common sense, not to speak of the theological heresies implied in the heliocentric view. They had made Aristotelian science their exclusive tool for understanding facts, just as today most laymen make their understanding of physical theory depend on their ability to visualize it in terms of simple mechanical models obeying Newtonian laws. But at the root of the tragic position of the Aristotelians was, in part, the fact that an acceptance of the Copernican theory as even a *possible* theory would have had to be preceded by a most far-reaching re-examination and re-evaluation of their personal beliefs. It would have required them to do the humanly almost impossible, to discard their common-sense ideas, to seek new bases for their old moral and theological doctrines, and to learn their science anew (which was of course what Galileo himself did to an amazing degree, for which his contemporaries called him fool, or worse, and for which we call him genius). Being satisfied with their system, the Aristotelians were, of course, unaware that history would soon prove their point of view to be far less effective in man's quest to understand nature.

Galileo's concrete observations meant little to them. The Florentine astronomer Francesco Sizzi (1611) argued in this manner why there could not, *must* not be any satellites around Jupiter:

"There are seven windows in the head, two nostrils, two ears, two eyes and a mouth; so in the heavens there are two favorable stars, two unpropitious, two luminaries, and Mercury alone undecided and indifferent. From which and many other similar phenomena of nature such as the seven metals, etc., which it were tedious to enumerate, we gather that the number of planets is necessarily seven . . . Besides, the Jews and other ancient nations, as well as modern Europeans, have adopted the division of the week into seven days, and have named them from the seven planets: now if we increase the number of planets, this whole system falls to the ground. . . . Moreover, the satellites are invisible

to the naked eye and therefore can have no influence on the earth and therefore would be useless and therefore do not exist."

A year after his discoveries, Galileo had to write to Kepler:

"You are the first and almost the only person who, after a cursory investigation, has given entire credit to my statements. . . . What do you say of the leading philosophers here to whom I have offered a thousand times of my own accord to show my studies, but who, with the lazy obstinacy of a serpent who has eaten his fill, have never consented to look at the planets, or moon, or telescope?"

Before this anecdote is again forgotten, let us re-emphasize its salient point, so forcefully expressed by Lodge (*op. cit.*):

"I know nothing of the views of any here present; but I have met educated persons who, while they might laugh at the men who refused to look through a telescope lest they should learn something they did not like, yet also themselves commit the very same folly. . . . I am constrained to say this much: Take heed lest some prophet, after having excited your indignation at the follies and bigotry of a bygone generation, does not turn upon you with the sentence '*Thou art the man.*'"

**10-2 Science and freedom.** The tragedy that descended on Galileo is described in many places, and it is impossible to do justice to the whole story without referring to the details. Briefly, he was warned in 1616 by the Inquisition to cease teaching the Copernican theory, for it was now held "contrary to Holy Scripture." At the same time Copernicus' book itself was placed on the Index Expurgatorius, and was suspended "until corrected." But Galileo could not suppress what he felt deeply to be the truth. Whereas Copernicus had still invoked Aristotelian doctrine to make his theory plausible, Galileo had reached the new point of view where he urged acceptance of the heliocentric system on its own merits of simplicity and usefulness, and apart from such questions as those of faith and salvation. This was the great break.

In 1623, Cardinal Barberini, formerly his dear friend, was elevated to the Papal throne, and Galileo seems to have considered it safe enough to write again on the controversial topic. In 1632, after making some required changes, Galileo obtained the Inquisitor's necessary consent to publish the work *Dialogue Concerning the Two Chief Systems of the World* (from which the previous arguments for the Copernican theory were drawn), setting forth most persuasively the

Copernican view in a thinly disguised discussion of the relative merits of the Ptolemaic and Copernican systems. After publication it was realized that he may have tried to circumvent the warning of 1616. Furthermore, Galileo's forthright and tactless behavior and the Inquisition's need to demonstrate its power over suspected heretics conspired to mark him for punishment.

Among the many other factors in this complex story a prominent role is to be assigned to Galileo's religious attitude—which he himself believed to be devoutly faithful, but which had come under suspicion by the Inquisition. Galileo's letters of 1613 and 1615 showed he held that God's mind contains all the natural laws, and that the occasional glimpses of these laws which the human investigator may laboriously achieve are proofs and direct revelations of the Deity quite as valid and grandiose as those recorded in the Bible. "From the Divine Word, the Sacred Scripture and Nature did both alike proceed. . . . Nor does God less admirably discover himself to us in Nature's actions than in the Scripture's sacred dictions." These opinions—held, incidentally, by many present-day scientists—can, however, be taken for symptoms of pantheism, one of the heresies for which Galileo's contemporary, Giordano Bruno, had been burned at the stake in 1600. Nor did Galileo help his case by such phrases as his quotation of Cardinal Baronius' saying: "The Holy Spirit intended to teach us in the Bible how to go to heaven, not how the heavens go."

The old, ailing man was called to Rome and confined for a few months. From the partly still secret proceedings we gather that he was tried (in absentia), threatened with torture, induced to making an elaborate formal renunciation of the Copernican theory, and finally sentenced to perpetual confinement. None of his friends in Italy dared to defend Galileo publicly. His book was placed on the Index (where it remained along with Copernicus' and one of Kepler's until 1835). In short, and this is the only point of interest to us here, he was set up as a warning for all men that the demand for spiritual obedience indisputably entails intellectual obedience also, that there can be no free science where there is no free conscience. His famous *Abjuration*, later ordered to be read from the pulpits throughout Italy and posted as a warning, reveals an ominous modern sound.\*

But without freedom, science cannot flourish for long. It is perhaps not simply a coincidence that after Galileo, Italy, the mother

\* The full condemnation and abjuration is printed as Appendix E in Sedgwick and Tyler, *op. cit.*



of outstanding men till then, produced for the next 200 years hardly a single great scientist, while elsewhere in Europe they arose in great numbers. To scientists today this famous facet in the story of planetary theories is not just an episode in passing. Not a few teachers and scientists in our time have had to face powerful enemies of open-minded inquiry and of free teaching, and again must stand up before those men who rightly fear the strength of unindoctrinated intellect.

Even Plato knew that an authoritarian state is threatened by intellectual nonconformists, and he recommended for them the now time-honored treatment: "re-education," prison, or death. Today we find Russian geneticists and astronomers being expected to reject well-established theories, not on grounds of persuasive new *scientific* evidence, but because of doctrinal conflicts. This same struggle explains the banishment from Nazi Germany's textbooks of the discussion of relativity theory because by the obscure standards of racist metaphysics Einstein's Jewish faith invalidated his work for Germans. This, too, was in part the meaning behind the "Monkey Trial" in Tennessee, where the teaching of Darwin's theory was to be suppressed for conflicting with certain types of Bible interpretations.

The warfare of authoritarianism against science, like the warfare of ignorance against knowledge, has not diminished since Galileo's days. Scientists take what comfort they can from the verdict of history, for less than 50 years later, Newton's great book, the *Principia*, had appeared, integrating the work of Copernicus, Kepler, and Galileo so brilliantly with the principles of mechanics that the long-delayed triumph of those pioneers of science was irrevocable, and more significant than they themselves might have hoped.

PROBLEM 10-2. In J. B. Conant's book, *On Understanding Science* [New Haven: Yale University Press (1947)] are summarized "Certain Principles of the Tactics and Strategy of Science" and "The Interaction of Science and Society." The following main points are made: (a) New concepts evolve from experiments or observations and are fruitful of new experiments and observations. (b) Significant observations are the result of "controlled" experiments or observations; the difficulties of experimentation must not be overlooked. (c) New techniques arise as a result of experimentation (and invention) and influence further experimentation. (d) An important aspect of scientific work is the interaction between science and society. Examine the development of planetary theory in terms of each of the four points (preferably after reading at least pages 101-109 in Conant's book).

### Further Reading

H. Butterfield, *op. cit.* (Ch. 2), Chapter IV.

W. S. Knickerbocker, *op. cit.* (Ch. 7). Chapter IV, excerpt from Galileo. Also H. Shapley and H. E. Howarth, *op. cit.*, pp. 41-57; and H. Boynton, *op. cit.*, pp. 27-43.

F. R. Moulton and J. J. Schifferes, *The Autobiography of Science*. Garden City: Doubleday, Doran and Co. (1945). Brief condensations. See pp. 63-77 (Galileo).

F. S. Taylor, *Galileo and the Freedom of Thought*. London: Watt and Co. (1938). An appraisal of Galileo's work and tribulations.

## CHAPTER 11

### NEWTON'S LAW OF UNIVERSAL GRAVITATION

**11-1 Science in the seventeenth century.** Between the time of Galileo's death and the publication of Isaac Newton's *Principia*\* lie barely 44 years, yet an amazing change of the intellectual climate in science has taken place during that time. On one side, the "New Philosophy" of experimental science is becoming a respected and respectable tool in the hands of vigorous and inventive investigators; and on the other, this new attitude is responsible for a gathering storm of inventions, discoveries, and theories. Even a very abbreviated list of [these, covering less than half of the seventeenth century and only the physical sciences, will show the justification for the term "the century of genius": the work on vacuums and pneumatics by Torricelli, Pascal, von Guericke, Boyle, and Mariotte; Wallis' work in mathematics; Descartes' great studies on analytical geometry and on optics; Huygens' work in astronomy and on centripetal force, his perfection of the pendulum clock, and his book on light; the improvement of microscopes and the invention of reflecting telescopes; Newton's work on optics, including the discovery of the solar spectrum, and his invention of calculus simultaneously with Leibniz; the opening of the famous Greenwich observatory; Roemer's measurement of the speed of light; and Hooke's work, including that on elasticity.

The men of science are no longer isolated. Although we must not suppose that suddenly *all* scientists are of a modern bent of mind, the adherents of the "New Philosophy" are many and growing in numbers. They have formed several scientific societies (Italy, England, France), of which the most famous is the Royal Society of London (as of 1662). They meet and communicate with one another freely now, they cooperate and debate, write copiously, and sometimes quarrel healthily. As a group, they solicit support for their work, combat attacks by antagonists, and have a widely read scientific journal. Science is clearly becoming well-defined, strong, and international.

---

\* Full title: *Philosophiæ Naturalis Principia Mathematica*; published in London, 1687.

What is behind this sudden blossoming? Even a partial answer would lead us far away from science itself, to an examination of the whole picture of quickening cultural, political, and economic changes in the sixteenth and seventeenth centuries.\* One aspect is that both craftsmen and men of leisure and money begin to turn to science, the one group for improvement of methods and products, the other for a new and exciting hobby. But availability of money and time, the need for science, and the presence of interest and organizations do not alone explain or sustain such a thriving enterprise. Even more important ingredients are able men, well-formulated problems, and good mathematical and experimental tools.

Able scientists—these were indeed at hand. Some, like Newton, were men in still primarily scholastic universities; others, like Newton's friends Edmond Halley and Christopher Wren, were initially gifted amateurs.

Well-formulated problems—these had recently become clear in the writings of Galileo. It was he, above all, who had directed attention to the fruitful language of science, who had presented a new way of viewing the world of fact and experiment. His scorn of sterile introspection and blind subservience to ancient dogma echoed through all fields of science.\*\* Plato's old question, "By the assumption of what uniform and ordered motions can the apparent motions of the planets be accounted for?" had lost its meaning in the new science; the new preoccupation is illustrated by what may be called the two most critical problems in 17th century science, "what forces act on the planets to explain their actually observed paths?" and "what accounts for the observed effects of terrestrial gravitation now that Aristotelian doctrines have failed us?"

Good mathematical and experimental tools were also being created. With mathematics finding expression in physics, the two fields cross-fertilized and gave rich harvests. The instrument maker and the scientist aided each other similarly. But in trying to solve the riddle of the spectacular rise of science after Galileo, perhaps the most important factor is the realization that science has an explosive type of growth once the necessary conditions are given.

---

\* Some of these problems are skilfully analyzed in G. N. Clark, *Science and Social Welfare in the Age of Newton*. Oxford University Press (1937 and 1949).

\*\* For a completely delightful example, read the first thirty or forty pages in the Everyman's Library edition of Robert Boyle's *Sceptical Chymist* (originally published 1661).

Around the time of Galileo the necessary conditions for growth were established: at last there were many men with similar attitudes working in the same fields; they could communicate more freely with one another and with the accumulated achievements of the past (partly through the relatively young art of printing); they had become impatient with qualitative reasoning and began to be more and more intrigued with the quantitative approach. To use a modern analogy, the critical mass was reached, and the chain reaction could proceed.

This, from the point of science, was the new age in which Newton lived. But a brief word of caution before we follow his work. The history of ideas, in science as in any other field, is not just an account of visits with the most important names. The work of each genius is made possible, is stabilized, is connected with the whole structure of science only through the labors of lesser-known men, just as a brick is surrounded by fortifying mortar. A house is not just a heap of bricks—and science cannot be made by giants only. As Lord Rutherford has said: "It is not in the nature of things for any one man to make a sudden violent discovery; science goes step by step, and every man depends on the work of his predecessors. . . . Scientists are not dependent on the ideas of a single man, but on the combined wisdom of thousands of men." Therefore, properly, we should trace in each man's contribution the heritage of the past, the influence of his contemporaries, and the meaning for his successors. It would be exciting and rewarding, but here we can touch only on the barest outlines.

**11-2 A short sketch of Newton's life.** Isaac Newton was born on Christmas day, 1642, in the small village of Woolsthorpe in Lincolnshire. He was a quiet farm boy who, like young Galileo, loved to build and tinker with mechanical gadgets and who seemed to have a secret liking for mathematics. Through the fortunate intervention of an uncle he was allowed to go to Trinity College of Cambridge University in 1661 (where he appears to have initially enrolled in the study of mathematics as applied to astrology!). He proved an eager and excellent student. By 1666, at twenty-four, he had quietly made spectacular discoveries in mathematics (binomial theorem, differential calculus), optics (theory of colors), and mechanics. Referring to this period, Newton once wrote:

"And the same year I began to think of gravity extending to the orb of the Moon, and . . . from Kepler's Rule [Third Law] . . . I deduced that the forces which keep the Planets in their

orbs must [be] reciprocally as the squares of their distances from the centers about which they revolve: and thereby compared the force requisite to keep the Moon in her orb with the force of gravity at the surface of the earth, and found them to answer pretty nearly. All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded Mathematicks and Philosophy more than at any time since."

From his descriptions we may conclude that during those plague years, having left Cambridge for the time to study in isolation at his home in Woolsthorpe, he had developed a clear idea of the first two laws of motion and of the formula for centripetal acceleration, although he did not announce the latter until many years after Huygens' equivalent statement.\*

After his return to Cambridge, he did such creditable work that he followed his teacher as professor of mathematics. He lectured, and contributed papers to the Royal Society, at first particularly on optics. His *Theory of Light and Colors*, when finally published, involved him in so long and bitter a controversy with rivals that the shy and introspective man even resolved not to publish anything else. As Bertrand Russell points out, "If he had encountered the sort of opposition with which Galileo had to contend, it is probable that he would never have published a line." Newton now concentrated most on an extension of his early efforts in celestial mechanics, the study of planetary motions as a problem of physics. In 1684 the devoted Halley came to ask his advice in a dispute with Wren and Robert Hooke as to the force which would have to act on a body executing motion along an ellipse in accord with Kepler's laws; Newton had some time ago found the rigorous solution to this problem ("and much other matter"). Halley persuaded his reluctant friend to publish the work which touched on one of the most debated and intriguing questions of the time. In less than two years of incredible labors the *Principia* was ready for the printer; its publication

---

\* This too must have been the time of the famous and disputed fall of the apple. One of the better authorities for this story is a biography of Newton written by his friend Stukeley in 1752, where we hear that on one occasion Stukeley was having tea with Newton in a garden under some apple trees, when Newton told him he realized "he was just in the same situation, as when formerly, the notion of gravitation came to his mind. It was occasion'd by the fall of an apple, as he sat in a contemplative mood . . ."

in 1687 established Newton almost at once as one of the greatest thinkers in history.

A few years afterward, Newton, who had always been in delicate health, appears to have had what we would now call a nervous breakdown. On recovering, and for the next 35 years until his death in 1727, he made no major new discoveries, but rounded out earlier studies on heat and optics, and turned more and more to writing on theology. During those years he received honors in abundance. In 1699 he was appointed Master of the Mint, partly because of his great interest and competent knowledge in matters concerning the chemistry of metals, and he seems to have helped in re-establishing the debased currency of the country. In 1689 and 1701 he represented his university in Parliament. He was knighted in 1705. From 1703 to his death he was president of the Royal Society. He was buried in Westminster Abbey.

**11-3 Newton's *Principia*.** In the original preface to his work—probably the greatest single book in the history of science—we find a clear outline:

“Since the ancients (as we are told by Pappus) esteemed the science of mechanics of greatest importance in the investigation of natural things, and the moderns, rejecting substantial forms and occult qualities, have endeavored to subject the phenomena of nature to the laws of mathematics, I have in this treatise cultivated mathematics as far as it related to philosophy [we would say ‘physical science’] . . . for the whole burden of philosophy seems to me to consist in this—from the phenomena of motions to investigate [induce] the forces of nature, and then from these forces to demonstrate [deduce] the other phenomena, and to this end the general propositions of the first and second Books are directed. In the third Book I give an example of this in the explication of the System of the World; for by the propositions mathematically demonstrated in the former Books, in the third I derive from the celestial phenomena the forces of gravity with which bodies tend to the sun and the several planets. Then from the forces, by other propositions which are also mathematical, I deduce the motions of the planets, the comets, the moon, and the sea [tides] . . .”

The work begins with a set of definitions—mass, momentum, inertia, force, centripetal force. Then follows a section on *absolute* and

relative space, time, and motion. To us it is of interest that Newton hardly anywhere made explicit use of the concept of *absolute* quantities, because, as he correctly states, "the parts of that immovable [absolute] space, in which those motions are performed, do by no means come under the observation of our senses."

Immediately thereafter we find the famous three Laws of Motion and the principles of composition of vectors (forces and velocities). At a later point follows an equally remarkable and important passage on "Rules of Reasoning in Philosophy." The four rules, reflecting the profound faith in the uniformity of all nature, are intended to guide scientists in making hypotheses, and in that function they are still up to date. The first has been called a *Principle of Parsimony*, the second and third, *Principles of Unity*. The fourth is a faith without which we could not use the processes of logic.

These four rules have been paraphrased as follows:

I. Nature is essentially simple, therefore we should not introduce more hypotheses than are sufficient and necessary for the explanation of observed facts. "Nature does nothing in vain, and more is in vain when less will serve." (This fundamental faith of all scientists is almost a paraphrase of Galileo's "Nature . . . does not that by many things, which may be done by few"—and he, in turn, perhaps with some irony, quoted the same opinion from Aristotle (see Section 10-1).

II. Hence, as far as possible, similar effects must be assigned to the same cause; e.g., "the descent of stones in Europe and in America; the reflection of light in the earth, and in the planets."

III. Properties common to all those bodies within reach of our experiments are to be assumed (even if only tentatively) as pertaining to all bodies in general; e.g., extension, mobility, inertia, the gravitational effect of the earth on the moon as well as on stone.

IV. Propositions in science obtained by wide induction are to be regarded as exactly or approximately true, until phenomena or experiments show that they may be corrected or are liable to exceptions.\*

PROBLEM 11-1. Compare these four rules regarding scientific hypotheses with criteria for a good theory discussed in Section 8-2, (3) through (5).

PROBLEM 11-2. As a review of Newton's laws of motion, find what is *wrong* with the following argument. Imagine a stone of mass  $m_s$  falling toward the earth (mass  $m_e$ ) with acceleration  $g$ . Evidently the force acting on the stone is the gravitational pull ( $F_{\text{grav}} = m_s g$ ) which, by Newton's third law, also acts on the earth with equal magnitude but in the opposite direction. As seen from the falling stone, the earth approaches the stone also with

\* Paraphrased after W. W. R. Ball, *Mathematical Gazette*, July 1914



acceleration  $g$ , and since the force on the earth is  $F_{\text{grav}}$ , we write  $F_{\text{grav}} = m_e g$ . Putting these equations together ( $F_{\text{grav}} = m_s g$ ;  $F_{\text{grav}} = m_e g$ ), we note that  $m_s g = m_e g$  or  $m_s = m_e$ ; i.e., *the stone must always have the same mass as the earth.* (Clue to answer: The equation  $F = ma$  has attached to it a long "text" that, among other things, specifies a system of coordinates with respect to which the acceleration is to be measured.)

**11-4 Derivation of the law of universal gravitation.** The three main sections of the *Principia* contain an overwhelming wealth of mathematical and physical discoveries; included are the proofs leading to the law of universal gravitation, proofs so rigidly patterned\* that it seems better to present here a derivation of this historic law in another plausible sequence. The arguments are sometimes rather subtle, and thus they afford a splendid illustration of the interplay between established laws, new hypotheses, experimental observations, and theoretical deductions in physical theory. The purpose of the following pages is to gain an understanding of the process, rather than to memorize the individual steps.

(a) Planets and satellites are not in equilibrium. If they were, if no net force were acting on them, they would be moving in straight lines instead of elliptical paths, in accord with Newton's first law of motion.

(b) Whatever the nature or the magnitude of the net force acting on a planet or on a satellite, its direction must be toward the center of motion at every instant. Newton derived this conclusion directly from Kepler's second law (*Principia*, Book I, Prop. I and II), and we may paraphrase his own argument as follows:\*\*

A body moves along a straight line at constant velocity; at equal time intervals  $\Delta t$  it traverses equal spaces, i.e., in Fig. 11-1(a),  $\overline{PQ} = \overline{QR} = \overline{RS}$ , etc. With respect to any fixed point  $O$ , the line from  $O$  to the moving body sweeps out equal areas during equal time intervals, for the triangles  $PQO$ ,  $QRO$ ,  $RSO$ , etc., are all of equal area, having equal bases and equal heights. Now imagine that this body experiences a brief and sudden blow at  $Q$ , a force directed exactly along  $QO$ . Of course, the direction of motion is modified. A velocity component has been added which, during time  $\Delta t$ , by itself would move the body from  $Q$  to  $Q'$  [Fig. 11-1(b)], but which, in conjunction with the original velocity toward  $R$ , adds up to a total displacement from  $Q$  to  $R'$ . But the area swept out during time  $\Delta t$

\* It is not without interest to note that the *form* of the arguments in Newton is along the traditional deductive pattern of Euclid, as, for that matter, was the work of St. Thomas Aquinas or Spinoza.

\*\* See also the similar rendition in Lodge, *op. cit.*, pp. 174-176.

is *not* affected by this reorientation! The area of  $OQR$ , which would have been covered in time  $\Delta t$  if no blow had been applied, is the same as the area of  $OQR'$  that actually was covered. (*Proof:*  $RR'$  is parallel to  $QQ'$ , thus  $\Delta OQR$  and  $\Delta OQR'$  have the same base and equal heights.) Therefore, area  $OQR'$  is also equal to area  $OPQ$ .

If no further force acts, the motion during equal time intervals  $\Delta t$  then proceeds from  $Q$  to  $R'$ , from  $R'$  to  $S'$ , etc. But a second blow at  $R'$ , again directed exactly to  $O$ , modifies the motion once more [Fig. 11-1(c)]. By the very same argument as before, we know that area  $OR'S'' = OR'S'$ . In general, we conclude that *centrally directed* forces applied at equal time intervals do not affect the areas swept out per unit time. There is no reason why those time intervals should not be chosen as we please, so that in the limit as  $\Delta t$  approaches zero, the centrally

directed force becomes a continuously acting centripetal force and the broken curve melts into a smooth one. Finally, we turn the argument around and, with Newton, say that since planets by Kepler's empirical second law do indeed sweep out equal areas per unit time, the force acting on each must be a continuous, centrally directed force. In the case of the ellipse, this center of force is one of the foci; for the circle, the center of the figure.

PROBLEM 11-3. Strictly speaking, we should also prove that blows *not* all directed to the same center  $O$ , but instead to any other points, will fail to yield an equal area law. Attempt such a proof.

(c) Now that we accept that the force must be centrally directed, a *centripetal* force, there appears the following crucial problem: "If a

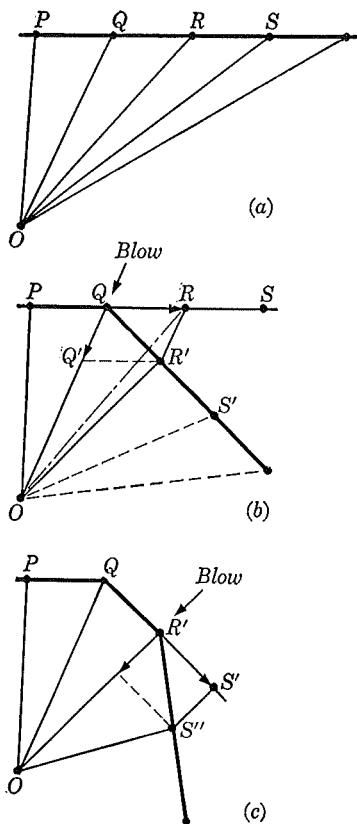


FIG. 11-1.

body revolves in an ellipse [including the special case of a circle], it is required to find the law of the centripetal force tending to the focus of the ellipse." Newton's mathematics proved for the first time that for paths along conic sections such as hyperbolas, parabolas, ellipses, and circles, the centripetal force at any instant must be proportional to the inverse square of the distance of the body to the focus. In short, any body obeying Kepler's first law of elliptical paths must be acted on by a force following the law  $F = C/R^2$ , where  $C$  is a constant for that body and  $R$  is measured from the center of the body to the center of forces, the focus of the ellipse.

We cannot at this point follow the general proof, but we will show that if for a planet in a *circular* path the centripetal force is granted to be equal to  $F = C/R^2$ , there follows by derivation, without further assumptions, that the celestial body should also obey the law

$$T^2 = K R^3.$$

Conversely, since we can observe the latter to be actually true—it is Kepler's third law—we judge that the hypothesis  $F = C/R^2$  is well founded for the case of planetary motion.

The derivation proceeds as follows. The centripetal force  $F_c$  (or now simply  $F$ ), on the planet which, as assumed, is given by  $C/R^2$ , is also by Newton's second law equal to  $m_p a_c$ , where  $m_p$  is the mass of the planet and  $a_c$  is the centripetal acceleration. For circular paths around the sun, as are actually followed by almost all planets with but little deviation,

$$a_c = \frac{v^2}{R},$$

$v$  being the speed of the planet in its orbit. But

$$a_c = \frac{v^2}{R} = \frac{4\pi^2 R^2}{T^2 R} = \frac{4\pi^2 R}{T^2},$$

where  $T$  is the period of orbital revolution of the planet. It follows [see Eq. (5-8)] that

$$F = m_p a_c = m_p \frac{4\pi^2 R}{T^2}. \quad (11-1)$$

Combining the last result with our assumed value for  $F$ ,

$$\frac{C}{R^2} = m_p \frac{4\pi^2 R}{T^2}, \quad \text{or} \quad T^2 = \left[ m_p \frac{4\pi^2}{C} \right] R^3. \quad (11-2)$$

Because  $m_p$  and  $C$  are constant, the bracketed term in Eq. (11-2) is constant for a given planet, no matter what the size of its orbit. Therefore  $T^2$  is proportional to  $R^3$ , as expected from Kepler's third

law. By yielding that law, our assumption  $F = C/R^2$  stands justified for the time being.

But we must recognize that this is only a tentative proof of the inverse square law for the centripetal force. Kepler's law requires

$$\left[ \frac{m_p 4\pi^2}{C} \right] = K,$$

where  $K$  is, for our solar system, the same constant for *all planets* in all their orbits around the sun. Not until we discover what  $C$  contains can we know whether the bracketed term actually yields the same constant for all planets. We shall strengthen and complete the above proof as soon as we can, for we have decided not to by-pass the difficult points. Let us note in passing that use has been made of Newton's second law of motion and of the equation for centripetal acceleration.

Historically, Newton's demonstration that elliptical planetary paths imply an inverse square law of the force came at a time when the idea of such a law was generally "in the air." In fact, Halley had come to Newton in 1684 just in order to ask whether he could supply the proof all others were looking for in vain!

(d) The *origin* of the centripetal force needed to keep the planets in their orbits has not been touched on so far. We recall that Kepler guessed that some driving magnetic force reached out from the sun to move the planets. He was wrong, but at least he was the first to regard the sun as the controlling mechanical agency behind planetary motion. Another picture had been given by the great French philosopher Descartes (1596-1650), who proposed that all space was filled with a subtle invisible fluid of contiguous material corpuseles; the planets of the solar system were supposed to be caught in a huge vortex-like motion of this fluid about the sun. This idea was attractive to the minds of the day, and consequently was widely accepted for a time, but Newton proved that this mechanism could not account for the quantitative observations on planetary motion as summarized, for example, in Kepler's laws. The problem remained.

At this point Newton proposed a dramatic solution: the centripetal force on the planets is nothing but a gravitational attraction of the sun, and the centripetal force on satellites revolving around planets is also completely given by the gravitational pull by the planets. (Less than a century before it would have been impious or "foolish" to suggest that terrestrial laws and forces regulated the whole universe, but now, after Kepler and Galileo had unified the

physics of heaven and earth, it had become a natural thing to suspect.) If the earth attracts the moon with the same type of force with which it attracts a falling apple or a projectile, and if the sun attracts earth, moon, and all other celestial bodies with the same type of force, then there is no need for any additional cosmic force or prime mover, then gravity becomes a universal, unifying principle which, while in fundamental contradiction to the axioms of the scholastics, would have gladdened the heart of Kepler.

But we must follow the proof. First paralleling young Newton's thoughts, let us see whether the centripetal force  $F$  needed to keep the moon in its (nearly enough) circular orbit about the earth *can* be identified with terrestrial gravity. The *direction* of  $F$  is by definition toward the center of the earth, and this checks with the direction of the force of gravity. But what about the *magnitude* of  $F$ ? We apply the equation for centripetal force to this case, and find

$$F = m_m \frac{4\pi^2 R}{T^2}, \quad (11-3)$$

where  $m_m$  is the mass of the moon,  $R$  is now its distance from the center of rotation about the earth, and  $T$  its period of revolution.

Does this value for  $F$  actually coincide with the gravitational pull which the earth exerts on our satellite, as Newton proposed? That depends on the nature of the gravitational force. If gravity is propagated undiminished through all space, the weight of the moon will be simply  $m_m \cdot g$ , the same as a stone of the moon's mass when placed on giant scales somewhere on the surface of the earth. But not only does it seem unlikely that the gravitational acceleration stays the same no matter how far away from the earth we go, we also recall that in part (c) above we found evidence that the centripetal force (whatever its final nature) must fall off as the square of the increasing distance. If gravity is to account completely for the centripetal force, it too will have to follow an inverse square law. Let us therefore assume that the weight of an object falls off according to such a law, and consider now whether the gravitational pull of the earth on the moon just equals the centripetal force in Eq. (11-3).

This is our argument: An object with the same mass as the moon,  $m_m$ , has the weight  $m_m g$  when weighed at the surface of the earth, i.e., at a distance  $r$  (the radius of the earth) from the earth's center. That same object, when removed to a great distance  $R$  from the earth's center, will assume a smaller weight, which we shall call  $W_R$ ,

and which must fulfill the following proportion if the inverse square law is to be obeyed:

$$\frac{m_m g}{W_R} = \frac{1/r^2}{1/R^2}$$

or 
$$W_R = m_m g \frac{r^2}{R^2}. \quad (11-4)$$

If the centripetal force  $F$  acting on the mass  $m_m$  rotating about the earth at distance  $R$  with a period  $T$  is really equivalent to the gravitational force  $W_R$  at that distance, the terms on the right side of Eqs. (11-3) and (11-4) should also be equivalent:

$$\frac{m_m 4\pi^2 R}{T^2} = m_m g \frac{r^2}{R^2}$$

or 
$$T^2 = \frac{4\pi^2}{g r^2} R^3. \quad (11-5)$$

Conversely, if we substitute observed values of  $T$ ,  $g$ ,  $r$ , and  $R$  in Eq. (11-5) and find that the equation yields equal numbers on each side, then we are justified in regarding our hypotheses valid; then the gravitational force *does* fall off as the square of the increasing distance, and *does* fully account for the needed centripetal force. [Incidentally, we note with satisfaction that Kepler's third law is implied in our result, Eq. (11-5).]

PROBLEM 11-4. Substitute the needed values in consistent units and check to what extent Eq. (11-5) holds. The period  $T$  of the moon is about  $27\frac{1}{3}$  days,  $g = 32$  ft/sec<sup>2</sup>,  $r = 3960$  miles,  $R = 60 r$ .

This was the calculation which Newton, with contemporary data, found to "answer pretty nearly," probably within a few percent. The assumption of a strictly circular path and somewhat inaccurate values of  $r$  and  $g$  made it clear from the start that no *perfect* agreement could have been expected.

It has been a source of much speculation why Newton did not tell anyone of this remarkable result when he first conceived it, or for nearly 20 years thereafter. Apart from his reticence and his fear of litigations with jealous men, he seems to have been unable at the time to account clearly for one implied assumption in the argument, namely, that the earth's gravitational force acts as if originating at the very center of the globe, and that consequently the measurement of distances must be made, not to the globe's surface, but to its center. He understood this requirement later when writing the

*Principia*, where he proved in general that two homogeneous spheres attract each other as though their masses were concentrated at the centers.

As a summary for part (d) we may use Newton's own statement "that the moon gravitates toward the earth, and by the force of gravity is continually drawn off from a rectilinear motion, and retained in its orbit." Apart from the assumptions and approximations alluded to, we have made use of only the following new arguments: that the gravitational force falls off as the square of the distance, and that by Rules I and II of Section 11-3 we can identify terrestrial gravity with the centripetal force on the moon.

(e) The previous paragraphs have not involved the forces between the sun and the planets, but we should again suspect that these ideas may be extended to the whole solar system. In Newton's words:

"The force which retains the celestial bodies in their orbits has been hitherto called centripetal force; but it being now made plain that it can be no other than a gravitational force, we shall hereafter call it gravity. For the cause of that centripetal force which retains the moon in its orbit will extend itself to all planets, by Rules I, II, and IV."

But such rules are just guiding, not prescriptive ones. They suggest, but do not prove; this labor remains to be done.

We remember that the centripetal force on planets in circular orbits around the sun was given by Eq. (11-1) as

$$F = m_p \frac{4\pi^2 R}{T^2}.$$

Does this actually also correspond in each case to the gravitational attraction of the sun on the particular planet? If we knew the value of  $g$  on the surface of the sun we could fashion the arguments as in the previous section; but of course we do not know it at this point, and so we have recourse to another argument. Perhaps we can decide on a theoretical formula for the gravitational force of any one body on another. With the confidence gained by our success in the previous discussion of the earth and moon, we now boldly suggest that the gravitational force  $F_{\text{grav}}$  (or, more briefly,  $F_G$ ) between *any* two spherically symmetric bodies is proportional to the inverse square of the distance between the two centers, everything else being kept constant:

$$F_G \propto \frac{1}{R^2}.$$

Next, let us consider two specific solid bodies, quite isolated from the rest of the universe, say a stone ( $m_1$ ) and the earth ( $m_2$ ) at a distance  $R$  between their centers. The pull of gravity, the weight of  $m_1$  at the given distance  $R$ , is  $F_G$ . But by Newton's law of action and reaction, the pull of the earth ( $m_2$ ) on  $m_1$  is just as large as the pull of  $m_1$  on  $m_2$ ; the weight of a stone  $F_G$  as measured by its attraction toward the earth is equal to the "weight of the earth" as measured by its attraction to the stone, strange though this may sound at first, and either one may be called  $F_G$  (Fig. 11-2).

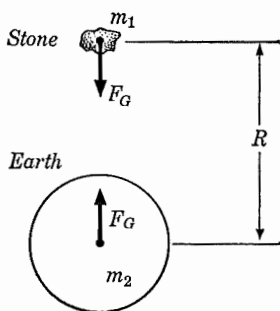


FIG. 11-2. The force of attraction of the earth on a stone is equal in magnitude to the attraction of the stone on the earth.

However, we know by experiment that at a given locality the weight of the stone grows directly with the mass of the stone, or  $F_G \propto m_1$ . On the other hand, if the mass of the planet changed, the weight of a given stone would change also (in fact, you would weigh much less on our little moon than on earth). In short, if experiments prove that  $F_G \propto m_1$  at a constant distance  $R$ , then we must also accept  $F_G \propto m_2$ , for otherwise we should have to make the assumption that the mutual gravitational pull  $F_G$  depends on something besides the magnitude of the masses and the distances between them.

Combining these three proportionalities,

$$F_G \propto \frac{1}{R^2}, \quad F_G \propto m_1, \quad F_G \propto m_2,$$

we have

$$F_G \propto \frac{m_1 m_2}{R^2}, \quad \text{or} \quad F_G = G \frac{m_1 m_2}{R^2}, \quad (11-6)$$

where  $R$  is the center distance between the two bodies (let us assume them to be homogeneous spheres) and  $G$  is a constant of proportionality.

If we have confidence that Eq. (11-6) gives, in fact, the correct attraction which two masses exert on each other, and if one were the sun ( $m_s$ ), the other any planet ( $m_p$ ), then the sun's pull would be

$$F_G = G \frac{m_p m_s}{R_{ps}^2}.$$



If the moon and the earth were taken separately, their mutual force would be

$$F_G = G \frac{m_m m_e}{R_{me}^2}.$$

We seem to have in Eq. (11-6) a *law of universal gravitation*.

(f) But now we begin to be assailed by doubts. If we are right, this newly found law must certainly be compatible with all three of Kepler's laws. There is no trouble about the first two—elliptical orbits and the equal area relations must indeed result from this type of gravitational force, since it is centrally directed and is proportional to the inverse square of the distances in accordance with just those requirements, laid down in parts (b) and (c). But what of Kepler's third law? We remember that our previous argument, centering on Eq. (11-2), was a little shaky. We had not really proved, with the force assumed there, that  $T^2/R^3$  had to be a truly universal constant, with the same value for *all* planets in our universe. Now that we know more about the force acting in the solar system, we can try more rigorously to obtain such a result; for if we do *not* obtain it, gravity cannot be the sole force among the celestial bodies.

Let us test our law of universal gravitation against the nearly circular orbital motion of our own planet. The centripetal force that must exist on our mass  $m_p$  at distance  $R_{ps}$  from the sun is given by

$$\frac{m_p 4\pi^2 R_{ps}}{T^2}.$$

The gravitational force supposedly available is  $G \frac{m_p m_s}{R_{ps}^2}$ . If the two coincide, if the gravitational force "supplies" and so "explains" the centripetal force, then

$$m_p \frac{4\pi^2 R_{ps}}{T^2} = G \frac{m_p m_s}{R_{ps}^2}, \quad \text{or} \quad T^2 = \left[ \frac{4\pi^2}{G m_s} \right] R_{ps}^3. \quad (11-7)$$

Now we really can see whether  $T^2/R_{ps}^3$  is a constant for all planets, as required by Kepler's third law. The brackets in Eq. (11-7), which is to be identified with that constant, contain only the constant of proportionality  $G$ , the mass of the sun, and a numerical factor; none of these is variable from planet to planet. Therefore  $T^2/R_{ps}^3$  is truly constant, Kepler's law is fulfilled, and the assumption of universal gravitation is vindicated.

(g) But there remains yet another doubt. Is it proper to have assumed just now so glibly that  $G$  does in fact have the same value

for all planets? (If it does not, the bracketed term above is, after all, not constant for all bodies.) Here we turn to experiment. We can first measure  $G$  only for various materials on earth; using the equation

$$F_G = G \frac{m_1 m_2}{R^2},$$

we get

$$G = F_G \frac{R^2}{m_1 m_2}.$$

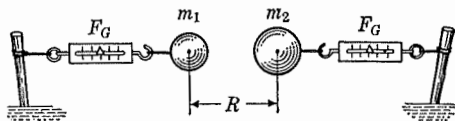


FIG. 11-3.  $F_G = G(m_1 m_2 / R^2)$ .

We may propose to measure  $F_G$  with spring balances, neglecting all other disturbing attractions except that between two known masses  $m_1$  and  $m_2$  at measured distance  $R$  from each other (see Fig. 11-3). It turns out that for any reasonably small masses in our laboratories,  $F_G$  is so exceedingly small that special techniques and precautions have to be adopted to measure any attraction between two such masses; for example, two 1-kgm masses separated by 10 cm pull on each other with a force smaller than 1/1000 dyne! The most serious technical problems of measurement were solved by Henry Cavendish over 100 years after the publication of the *Principia*, and the best present value for  $G$  is about  $6.67 \times 10^{-8}$  dyne-cm<sup>2</sup>/gm<sup>2</sup> for all substances. (Cavendish's measurements were made with a delicate torsion balance.)

**PROBLEM 11-5.** If a remeasurement of  $G$  were attempted with the arrangement shown in Fig. 11-3, and if the error inherent in our measuring devices is about 20 dynes, what is the smallest usable radius for each of two similar gold spheres which almost touch? (Density of gold = 19.3 gm/cm<sup>3</sup>. This is a fairly representative problem of what one has to go through before designing apparatus for experimentation.)

**PROBLEM 11-6.** If by chance our standard of mass were not 1 gm, but some other unit 3880 times larger (call it 1 tm), what would be the value of  $G$  in that other system of units? Does this result seem to teach any particular lesson?

Cavendish's results, and all the more accurate ones since, have shown that  $G$  has indeed the same value, no matter what the composition of  $m_1$  and  $m_2$ , even for meteorite material. Being disposed to apply terrestrial laws to celestial bodies, we extend our findings, and maintain that in the absence of evidence to the contrary all materials in the world, including sun, planets, and satellites, obey the same law of gravitation.

Although Newton knew in principle how to go about measuring  $G$

accurately, he lacked the precise instruments; nevertheless he devised an ingenious proof for the *constancy* of  $G$ , which is the main point at issue here. Consider a mass  $m_1$  on the surface of our earth (of mass  $m_e$ ), i.e., at a distance  $r$  from the center of our globe. Its weight, which we called  $F_G$ , is of course also given by  $m_1g$ , therefore, by our new law,

$$m_1g = G \frac{m_1m_e}{r^2}, \quad \text{or} \quad G = \frac{r^2}{m_e} g. \quad (11-8)$$

At a given locality,  $r^2/m_e$  is, of course, a constant, regardless of its numerical value. If at that locality all substances show precisely the same value for  $g$  (the gravitational acceleration), then we have established that  $G$ , too, is constant, regardless of chemical composition, texture, shape, etc. This is just what Newton showed experimentally. His measurements of  $g$  were made not by just dropping small and large bodies, from which even Galileo had concluded that  $g$  cannot vary significantly, but by the much more accurate method of timing pendulums of equal lengths but of different materials. It was known by then that for a given length the period  $T$  of a simple pendulum is proportional only to  $1/\sqrt{g}$ . After exhaustive experiments, all pointing to the constancy of  $G$ , he could write: "This is the quality of all bodies within the reach of our experiments; and therefore (by Rule III) to be affirmed of all bodies whatsoever."

Thus  $G$  attains the status of the *universal* constant of gravitation (one of the very few truly universal constants in nature), and the proposed law of universal gravitation in Eq. (11-6) can be applied on a cosmic scale. And quite incidentally, Eq. (11-8) clears up a question on which both ancient Aristotelians and present-day students harbor strongest suspicions; namely, the reason why at a given locality the acceleration of gravity is constant for all bodies. The answer is that  $g$  depends only on  $G$ ,  $m_e$ , and  $r^2$ , all constant at one locality.

(h) Once a value for  $G$  is at hand, we may determine the mass of the earth [from Eq. (11-8)] and of the sun [from Eq. (11-7)].

Moreover, Eq. (11-7) should apply in the same form for a satellite moving around a planet of mass  $m_p$ , with a period  $T_{\text{sat}}$  and a radius (or a half major axis of the elliptical orbit)  $R_{p\text{-sat}}$ :

$$T_{\text{sat}}^2 = \frac{4\pi^2}{Gm_p} R_{p\text{-sat}}^3. \quad (11-9)$$

This yields the mass of any planet whose satellites we can observe.

PROBLEM 11-7. Determine the mass of the earth and of the sun as indicated. (Answer:  $m_{\text{earth}} = 5.98 \times 10^{27}$  gm,  $m_{\text{sun}} = 333,000 \times m_{\text{earth}}$ .)

PROBLEM 11-8. The innermost of Saturn's nine satellites, Mimas, has a fairly circular orbit of 187,000 km radius and a period of revolution of about 23 hours. Find the mass of Saturn.

For the masses of the satellites themselves (including our own moon), and of the planets which boast of no moons (Mercury, Venus), no such simple calculations exist. There we may use the relatively minor and elaborate interactions between the celestial bodies, which we have conveniently neglected until now. Some of these are called *perturbations*, i.e., slight digressions from the regular path of a body because of the pull of other celestial objects. Another detail which may be used to derive the relative mass of a satellite is the fact that Kepler's laws hold, strictly speaking, only if the center of mass of the system is considered to be at the focus of the ellipse. Furthermore, it is not the earth which moves in accordance with the first law in an elliptical orbit around the sun, but the center of mass of earth and moon together (a point about 1000 miles below the surface of the earth and on a line between the centers of those two spheres).

Except for such mass determinations (or for accurate astronomical and nautical tables), we usually forget about the complications. But one of the results obtained is that the moon's mass is about  $\frac{1}{82}$  of the earth's, and that the sun's mass, determined in this indirect fashion, has the same value as followed from Eq. (11-7).

It follows also that, owing to perturbations, clearly Kepler's laws cannot be expected to hold *exactly*. (In fact, astronomers before Newton had begun to worry more and more about these discrepancies, particularly for Saturn and for the moon.) Once more we find a set of laws breaking down when the precision of observations is increased beyond the original range. But this breakdown is not serious; it does not invalidate the over-all outlines of the theory of planetary motion, particularly now that it is possible in principle to account in detail for each discrepancy by means of the law of gravitation.\*

As a summary for part (h), we may state that the law of gravita-

---

\* We are reminded of the 8 minutes of arc which Copernicus did not know of, and are tempted to marvel again how significant it is that these secondary effects were too small to appear earlier and confuse the original investigators. But in this case there seems to be an extra twist; if the perturbations were any larger, we might well not have a solar system to contemplate, because it would hardly be stable enough to last these billions of years without catastrophic collisions.

tion, which we formulated originally by arguments for terrestrial bodies, was found to hold also among the planets, because by its assumption Newton could predict relationships (such as Kepler's third law) well established by observation. Furthermore, the law opened up ways of calculating the mass of the sun, the planets, and the satellites. Some approximate relative masses are given in Table 11-1.

Table 11-1. Mass relative to earth ( $1.00 = 5.98 \times 10^{27}$  gm)

Sun	333,000	Jupiter	318
Moon	1/81.3	Saturn	95.3
Mercury	0.045	Uranus	14.6
Venus	0.82	Neptune	17.3
Mars	0.108	Pluto	0.1 (?)

PROBLEM 11-9. With the data given in Table 11-1, compare the gravitational pulls of sun and earth on the moon. Why does the earth not lose the moon to the sun? Can the path of the moon ever be convex as seen from the sun? [For hints you may wish to read F. L. Whipple, *Earth, Moon and Planets*, p. 30, Philadelphia: The Blakiston Co. (1941)].

PROBLEM 11-10. Equation (11-7) can be written as

$$K = \frac{4\pi^2}{Gm_s},$$

where  $K$  is the constant in Kepler's third law as obtained by observation of  $T$  and  $R$ . You have taken this equation to find  $m_s$ , using also the measured value of the universal constant  $G$ . The value of  $m_s$  so obtained actually checks against independent calculations of  $m_s$  using perturbation effects. Exactly in what way is our whole argument strengthened by this gratifying confirmation?

PROBLEM 11-11. (a) Using the calculated values of the masses for earth, moon, and sun, the given distances of sun and moon from earth, and the fact that the angle from our eye to either edge of both sun and moon is about  $\frac{1}{2}^\circ$ , find the density of the earth, moon, and sun. (The density of most rocks is about 3 gm/cm<sup>3</sup>.) (b) From Eq. (11-8) find the acceleration of gravity on the moon and the sun.

PROBLEM 11-12. An artificial satellite or floating observation platform is to be shot up to revolve around the earth like a little moon. What must be its period and speed of revolution around the center of the earth if it is to stay at a height of 1000 miles from the surface? Outline qualitatively the requirements for the initial velocity if this satellite is to be shot from a big gun on the earth.

(i) We have extended our law of universal gravitation in *one* direction—to all planets and satellites; it seems logical to extend it in the other direction—to all *parts* of every body. Newton writes in Book III of the *Principia*:

“Proposition VII. Theorem VII: That there is a power of gravity tending to all bodies, proportional to the several quantities of matter [i.e., the product of the masses] they contain. That all the planets mutually gravitate one towards another, we have proved before; . . . [we now introduce the idea that] the force of gravity towards any whole planet arises from, and is compounded of, the forces of gravity toward all its parts . . . If it be objected, that, according to this law, *all* bodies with us must mutually gravitate one toward another, whereas no such gravitation anywhere appears, I answer, that . . . the gravitation toward them must be far less than to fall under the observation of our senses.”

With the aid of his own calculus Newton proceeds to show that assuming the same universal law of gravitation for each smallest particle of a body (say of a sphere), we obtain in sum a resultant force of gravity for the whole body of such characteristics as are actually observable.

There offers itself now an over-all view of the enormous range and sweep of the simple formula for the gravitational attraction of two bodies. *Postulating* the law just for each of the parts of a body, it gives us the attraction of the whole for some outside object. For the special case of the spherical celestial bodies, it provides a force necessary and sufficient for the description of all observed motions, for the derivation of all three laws of Kepler, and for the small, long-observed deviations therefrom. This was the vindication of Copernicus, Kepler, and Galileo—in the context of the greater, more far-reaching theory of universal gravitation. The whole structure is often referred to as the *Newtonian Synthesis*.

PROBLEM 11-13. List the fundamental hypotheses of the theory of universal gravitation and the experimental consequences. Which hypothesis is not *directly* confirmable by experiment (although its consequences are)?

11-5 **Some influences on Newton's work.** What were the main intellectual tools, concepts, and attitudes with which Newton worked?

If we wish to claim any insight into his time and his work we must at least summarize these facts now; most of them have already appeared scattered here and there through the discussion.

(a) Newton was not a "hardboiled" scientist of the 20th century-type, but was a child of his time. He was not free from all traces of what we regard as pseudosciences; apart from some early interest in astrology, he seems to have spent much time in his "elaboratory," cooking potions that to us would smell more of alchemy than of chemistry—although the aim there, as in all his activities, seems to have been the search for underlying general principles, not, of course, quick practical gains.\* By our present standards we might also say that his belief in absolutes and his anthropomorphic conception of a Creator carried very far into some of his scientific writings. But here we touch on the very secret of motivation for his scientific work; such a study, although fascinating in its own right, must to us remain a detail. First of all, we should consider the clear Galilean influence on Newton's formulation of the physical concepts of mass and the like. The decisive attitude throughout his work is that celestial phenomena are explainable by quantitative terrestrial laws; that these laws have legitimate general meaning and are not just mathematical conveniences covering unattainable "true" laws.

(b) To his fundamental faith in the proximity and accessibility of natural law, we must add a word about his methodology. His debt to the pioneers of the new experimental science is clear (for example, he made ingenious pieces of equipment and performed good experiments when the theory needed verification), but he also successfully combined this principally inductive approach with the deductive method then most prominently displayed in Descartes. With his mathematical powers enriching the experimental attitude, he set a clear, straight course for the methods of physical science.

(c) Not just his attitude towards concepts, but many of the concepts themselves came to him from Galileo and his followers, above all, force, acceleration, the addition of vector quantities, and the first law of motion. Newton also drew, of course, from Kepler; and through their books and reports to the Royal Society he kept in touch with the work of such contemporaries as Huygens and Hooke.

(d) Apart from his own experiments, he took his data from a great variety of sources; for example, Tycho Brahe was one of several

\* Sidelights on Newton's work as experimenter, chemist, theologian, and King's servant can be found in the collection of the History of Science Society, *Sir Isaac Newton (1727-1927)*, Williams & Wilkins Co., Baltimore, 1928.

astronomers, old and new, whose observations of the moon's motion he consulted. When he could not carry out his own measurements he knew whom to ask, and we find him corresponding widely with men like Flamsteed and Halley, both Royal Astronomers. There are evidences that he searched the literature very carefully when he was in need of exact data, for example, on the radius of the earth and the distance to the moon.

(e) Lastly, we must not fail to consider how fruitfully and exhaustively his own specific contributions were used repeatedly throughout his work. The laws of motion and the mathematical inventions appear again and again, at every turn in the development. But he was modest about this part of his achievement, and said once that if he had seen further than others "*it is by standing upon ye shoulders of Giants.*"

**11-6 Some consequences of the law of universal gravitation.** What amazed Newton's contemporaries and increases our own admiration for him was not only the range and genius of his work on mechanics, not only the originality and elegance of his proofs, but also the detail with which he developed each idea to its fullest fruition. It took almost a century for science to fully comprehend, verify, and round out his work, and at the end of a second century an important scientist and philosopher still had to confess: "Since his time no essentially new principle [in mechanics] has been stated. All that has been accomplished in mechanics since his day has been a deductive, formal, and mathematical development of mechanics on the basis of Newton's laws."

(a) An example of Newton's thorough mastery was his treatment of the moon's perturbations. The influence on its path of the gravitational forces from almost all other celestial bodies was considered, and the repeated use of the gravitational law yielded an astonishingly close approximation to every small complexity of motion. (A full list of influential factors in tabular form would now fill hundreds of pages.) For one minor variation of the path, however, Newton's published theoretical result was only half the observed value. He did not camouflage this defect, but clearly stated that his calculations gave the wrong figure. Consequently, a loud, long battle arose among some scientists, several of whom called for complete abandonment of the whole theory of gravitation on account of this discrepancy. Eventually the mathematician Clairaut noted a small error in the long calculations, clearing up the difficulty. Later still, some unpublished



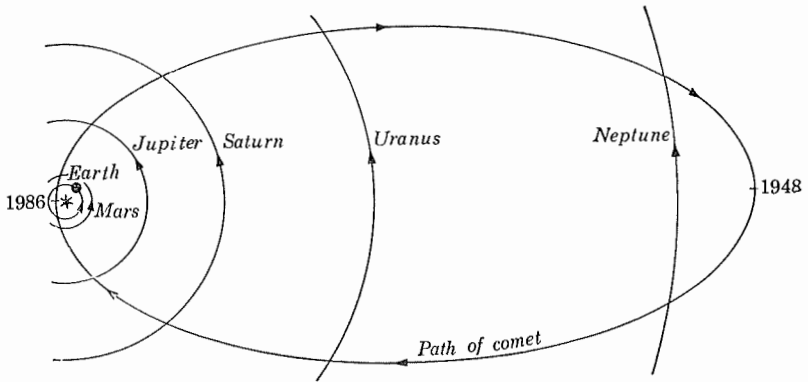


FIG. 11-4. Path of Halley's Comet.

notes of Newton were found, showing that he, too, 50 years before Clairaut, had discovered and corrected his own mistake!

(b) Comets, whose dreaded appearance had been interpreted as sure signs of disaster throughout antiquity and the middle ages, were now shown to be nothing but passing clouds of material obeying the laws of gravitation, made visible by reflected light while near the sun. The great comet of 1682, whose path Edmond Halley watched carefully, indicated to him a period of recurrence of *approximately* every 75 years if the comet, although an eccentric member of the solar family, nevertheless were to obey the usual laws of mechanics, including Kepler's laws. Its return in 1756 and twice since, after covering a wide ellipse carrying it far beyond the last planet, were heralded as significant symbols of the triumph of Newtonian science. (See Fig. 11-4.)

(c) Newton reasoned that the shape of planets and satellites might be explained as the result of the mutual gravity of the separate parts, which would pull a large quantity of initially liquid material (or a dust cloud) into a compact sphere. Further, although a motionless body might indeed form a perfect sphere, a planet rotating about an axis ought to assume the shape of an oblate spheroid, that is, it ought to bulge at the equator and be flatter at the poles. In fact, from the relative magnitude of the equatorial bulge, say of Jupiter as seen through a good telescope (see Fig. 11-5), one may calculate the period of rotation about the axis (the length of a day on Jupiter).

For the earth, Newton predicted that the same effect would be found, and later, accurate measurements actually revealed the diameter of the earth from pole to pole to be less by 27 miles than across

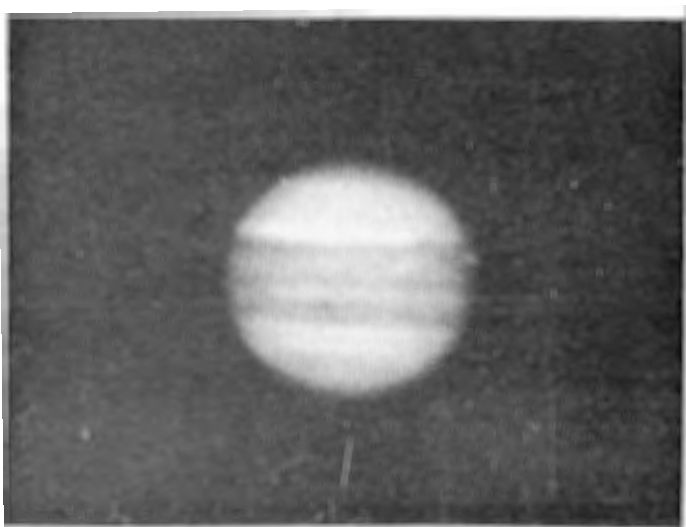


FIG. 11-5. The planet Jupiter.

the equatorial plane. Consequently the gravitational force on an object at the pole is somewhat greater than at the equator, owing to the shorter distance from the center of gravity of the earth to the pole. This effect is relatively small, but another cause joins to diminish the acceleration of a freely falling body at the equator. At that region the surface velocity of a point on earth is about 1000 mi/hr, due to the rotation of the globe about its axis. Since the centripetal acceleration  $a_c$  is given by  $v^2/r$ ,  $a_c$  for any object at the equator is about  $\frac{1}{16}$  ft/sec<sup>2</sup> or 3 cm/sec<sup>2</sup>, as we have calculated in Chapter 5. This much acceleration is required just to keep the body from flying off along a tangent; consequently the acceleration of free fall (and therefore the apparent weight of the body) is proportionally less than it would be if our globe were not rotating. Measured values of the free-fall acceleration, all at sea level, range from about 983 cm/sec<sup>2</sup> near the poles to 978 cm/sec<sup>2</sup> near the equator (cf. Chapter 4) with 980.398 cm/sec<sup>2</sup> the standard value for Cambridge, Massachusetts. These observed values are, of course, the true acceleration due to

gravity *minus* the centripetal acceleration at that latitude. But we shall be content with the convention that the letter  $g$  may stand for those *observed* values and yet be usually called "acceleration due to gravity."

PROBLEM 11-14. Because of the bulging of the earth near the equator, the source of the Mississippi River, although high above sea level, is nearer to the center of the earth than is its mouth. How can the river flow "uphill"? (*Hint*: The word "uphill" needs contemplation.)

PROBLEM 11-15. According to speculative theories, our moon might be material which was pulled out from the earth by a passing celestial body or which was shaken off by the rotating earth. How fast would the earth have had to rotate at that time to make the latter picture plausible?

PROBLEM 11-16. What approximate percentage error does the quoted value for  $g$  in Cambridge imply? How might such an accurate experiment be performed?

At present, the methods of measurement have been refined to such an extent that small local fluctuations in  $g$  can be detected and used to give clues of possible mineral deposits underground. But once one goes below the surface into a mine,  $g$  is found to change, becoming zero as the center of the earth is approached. The law of gravitation in its simple form holds only if the two bodies whose mutual attraction is to be measured are separated and not entwined or intergrown. Newton predicted this excellently, and a moment's thought will persuade us that at the center of the earth, at any rate, we could have no weight because the forces of the earth's attraction from all sides would cancel.

(d) The phenomenon of the tides, so important to navigators, tradesmen, and explorers through the ages, had remained a mystery despite the efforts of such men as Galileo, but at least the main features, the diurnal high tide and the semimonthly spring tide (maximum high tide), were explained by Newton through application of the law of universal gravitation. He recognized that the moon (and to a lesser extent the other celestial bodies) would pull on the nearest part of the ocean, and so tend to "heap up" the waters. We do not want to take time for the many details involved, except to look at the conclusions. The bulge of water, which actually is raised at the same time on *both* sides of the globe, does not stay directly under the moon but is displaced to a position a little ahead by the rotation of the earth. Thus a high tide at a point in the ocean is encountered some time after seeing the moon pass overhead, and again about 12 hours later (Fig. 11-6). The sun has a similar though smaller

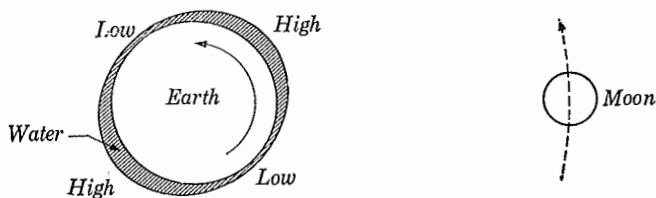


FIG. 11-6. Tides caused by the moon. The earth is to be imagined as rotating in this (much exaggerated) envelope of water. The high tide stays relatively fixed with respect to the moon.

effect,\* and since sun, moon, and earth are in line twice a month (at new and full moon), the two tidal forces coincide to produce semi-monthly extremes of tide level. Between spring tides, sun and moon may partially cancel each other's effects (neap tides).

The exact features of this complex phenomenon depend to a great extent on the topography of shore line and ocean floor. Long series of actual observations help to predict individual features of the tides at each locality. In passing, we should notice that the "solid" earth has similarly explained tides, which must be allowed for in such exact experiments as refined astronomical observations, and that it is just such a tidal effect that offers one of the methods for the approximate calculation of the mass of the moon.

(e) One of the most intriguing aspects of Newton's work still remains to be told: how, over 100 years after his death, Newton's law of universal gravitation helped to discover new planets. Lodge refers to it as follows:

"The explanation by Newton of the observed facts of the motions of the moon, the way he accounted for precession and nutation and for the tides, the way in which Laplace [whose mathematical work extended Newton's calculations] explained every detail of the planetary motions—these achievements may seem to the professional astronomer equally, if not more, striking and wonderful; . . . But to predict in the solitude of the study, with no weapons other than pen, ink, and paper, an unknown and enormously distant world, to calculate its orbit when as yet it had never been seen, and to be able to say to a practical astronomer, 'Point your telescope in such a direction at such a time, and you will see a new planet hitherto unknown to man'—this must always appeal to the imagination with dramatic intensity . . ."

\* The direct pull of the sun on the earth is about 175 times that of the moon, but tides are generated by the *difference* of pull on the waters on both sides of the globe; and that is larger for the moon. (Why?)

One night in 1781 William Herschel of Bath, England, an extraordinarily energetic mixture of professional musician and gifted amateur astronomer, was searching the sky with his homemade, 10-foot telescope. For years he had patiently looked at and re-examined every corner of the heavens, finding new stars, nebulae, and comets, and he was becoming well known among astronomers. On that particular night he again found a celestial object, hitherto uncatalogued and of such "uncommon appearance" that he suspected it to be a new comet. Through the Royal Society the news spread. As observations continued night after night, it became evident that this was not a comet but a newly discovered planet, over 100 times larger than the earth and nearly twice as far from the sun as Saturn, until then the outermost member of the solar family. So *Uranus* was discovered—an unsuspected and sensational widening of the ancient horizon, a planet just barely visible to the naked eye, but previously mistaken for a star when occasionally observed.

By that time it was known how to compute the elliptic orbit of a planet from a few widely separated observations of its varying positions. Also, the expected small deviations from the true ellipse owing to the perturbing force of the other planets were accurately predictable on the basis of Newton's law of gravitation. *Uranus'* 84-year orbit was so mapped out for it by calculation, and all went well for many years. But by 1830 it became more and more evident that *Uranus* was misbehaving, and that the assumptions on which its schedule had been worked out were in need of revision.

Some thought that Newton's theory might not apply accurately, after all, over such immense distances, but they had nothing better to offer (and as has been well pointed out, theories are in general not overthrown by isolated contradictory facts but only by other, better theories). Others suggested that a hitherto unsuspected comet or a more distant planet might introduce additional perturbations into *Uranus'* path, but they, too, were merely guessing and made no concrete quantitative predictions.

But the idea of an undiscovered planet beyond *Uranus* intrigued John C. Adams, a young undergraduate at Cambridge University, and he undertook the immensely difficult mathematical task of locating the positions of this suspected perturbing body solely from the observed motions of *Uranus*, using throughout the law of gravitation in unmodified form. He had his mathematical result of the problem two years after graduation, and for the necessary confirmation wrote to the Royal Observatory at Greenwich, asking that their powerful telescope search for the hypothetical new planet at a predicted loca-

tion beyond Uranus. Since Adams was an unknown young mathematician, however, he was not taken seriously enough to interrupt the current work there.

A few months later, another young man, Leverrier in France, published the result of similar, independent calculations, giving very nearly the same position for the suspected planet as had Adams. While finally some observations were slowly being undertaken in England to verify Adams' theoretical conclusions, Leverrier sent his own prediction to the head of the observatory at Berlin, who on the very evening of the letter's arrival himself searched for and recognized the planet at very nearly its predicted position. Thus *Neptune* was added to the solar system in 1846; and here was indeed a triumph of the law of universal gravitation!

Neptune, in turn, was closely watched. The orbital radius is about 30 times the earth's, and its period is therefore given by Kepler's law (correctly) as 164.8 years. But in time the perturbations of this planet and those of Uranus again became larger than was accountable in terms of known forces, and the hypothesis of still another planet was naturally made. An arduous 25-year long search yielded the discovery of *Pluto* in 1930, announced on the double anniversary of Herschel's discovery of Uranus and of the birthday of Percival Lowell, whose calculations had led to the search and who had founded the observatory in Arizona at which the discovery was made. Another astronomer, W. H. Pickering, had made independent calculations and predictions of Pluto's position as far back as 1909, and had initiated a telescopic search for the planet at Mount Wilson Observatory in California. Nothing was found then, but after the discovery at the Lowell Observatory in 1930, the old Mount Wilson photographs were re-examined and showed that Pluto could have been found in 1919 if its image had not fallen directly on a small flaw in the photographic emulsion!

This story dramatizes the frequently forgotten possibility that for every overpublicized discovery which is made by "accident," without elaborate preparations, there may well be an equally important discovery which, also by accident, *failed* to be made despite careful research.

(f) This chapter would be incomplete without a short discussion of the influence of a strange and simple rule on the discovery of Neptune. This was known as *Bode's law*, after the man who published it in 1772.

At the start of his long search for regularity in the solar system,

Kepler had discovered a crude relationship between the sizes of the various planetary orbits, but it was a rather fruitless jumble of numerological statements. The best that can be said of it was that its mathematical ingenuity first brought him to the attention of Brahe and Galileo. Later, the three Keplerian laws, unified by Newton's work, showed that there existed a simple universal relation between each planet's speed and orbit, but it left unanswered the question why a given planet did not move in some other possible orbit with a correspondingly different speed. Was there, for example, any specific reason why the earth could not pursue a course nearer to Mars', although necessarily with a relatively slower period of revolution? Nothing in Kepler's or Newton's work showed that the chosen orbit was in any way unique or necessary.

Since the whole effort of science is bent on finding simple regularities, it is no wonder that men continued to look for clues. Bode's law, obtained no doubt, like Kepler's laws, by incessant mathematical play with the data to be explained, established just such a clue, which, in the absence of physical reasons, must certainly look like a strange coincidence.

If we number the successively more distant planets from the sun with  $n = 1$  (Mercury), 2 (Venus), 3 (Earth), . . . , then the radius of the orbit of any planet (in astronomical units) is given by the *empirical* formula

$$R = 0.3 \times 2^{(n-2)} + 0.4 \text{ (in A.U.)}.$$

But this formula works only if two exceptions are made. One is that a zero must be used instead of the factors  $0.3 \times 2^{(n-2)}$  for the calculation of Mercury's orbital radius. Furthermore, although Mars can be assigned  $n = 4$ , Jupiter, the next planet, has to be assigned  $n = 6$ ; Saturn, the next and outermost planet then known, has to have  $n = 7$ . The number  $n = 5$  could not be given to any existing planet. This at once suggested to Bode that the disproportionately large space which actually extends between Mars and Jupiter harbored a yet unnoticed planet for which  $n = 5$  and which, *according to this rule*, could be expected to have an orbital radius of 2.8 A.U. (Derive this predicted radius from the last formula.)

Here was a chance to confirm whether the rule was anything more than coincidence and fantasy! Bode was convinced that a search would reveal the suspected planet, and he wrote, in a manner which reminds us again of Copernicus' and Kepler's motivations, "From Mars outward there follows a space . . . in which, up to now, no

planet has been seen. Can we believe that the Creator of the world has left this space empty? Certainly Not!"

At first no one discovered anything in that seemingly Godforsaken gap, and interest in the law must have lagged (one might say it did not fulfill the criterion of stimulation of further discoveries). But then, nine years after Bode's announcement, came Herschel's discovery of distant Uranus. With  $n = 8$  as a logical number to assign, the law predicted an orbital radius of 19.6 A.U., only about 2% larger than the actual value derived by subsequent observation.

This agreement directed attention to the other predictions inherent in Bode's law, and the search for the "missing" planet was renewed with vigor. In time, the first result came, but, as so often in science, only as the by-product of research not directly concerned with this problem. In 1801, the Sicilian astronomer Piazzi, while compiling a new star catalogue, chanced on a new celestial object which was moving as rapidly as might a comet or a planet. In the exciting months that followed, astronomers and mathematicians cooperated to derive the orbit of the foundling, now christened *Ceres*. It was disappointingly puny—less than 500 miles in diameter—but the orbital radius came to 2.77 A.U., only about 1% less than Bode's law predicted for the missing planet in that region between Mars and Jupiter!

Even while *Ceres* was being hailed as the long-sought sister planet, another, smaller one was discovered with about the same orbit, and then others still, all smaller than *Ceres*. Today we know of almost 2000, some with very eccentric paths, but all evidently springing from one family with a common history—perhaps the fragments which were forming into one planet when its evolution was somehow interrupted, perhaps the shattered remnants of one larger planet in the predicted orbit. It is possible to speculate that there were two (or more) planets trying to establish orbits near each other, but that they eventually collided "because" Bode's law would permit only one orbit to be there. This line of reasoning would put Bode's law into a new light; we might be tempted to search later for other clues that the law expresses simply the series of dynamically possible, stable orbits for single planets.

The discovery of these *asteroids* or *planetoids* bore out Bode's law so strikingly that it was natural to apply it when the suspicion of a planet beyond Uranus first arose. The first available value for  $n$  being 9, the corresponding orbital radius would be

$$(0.3 \times 2^{(9-2)} + 0.4) \text{ A.U.},$$



that is, about 38.8 A.U. Adams and Leverrier based their calculations on the predictions of Bode's law to find the probable location, mass, and orbit of their hypothetical planet. When Neptune was afterwards seen and its course plotted, the actual orbit was found to be about 20% smaller than Bode's law had predicted—the first deficiency in the law, but by luck one not so serious as to invalidate the calculations that led to Neptune's discovery. Later, when Pluto was found, its actual orbital radius of 39.46 A.U. much more seriously contradicted the value of 77.2 A.U. which would follow from Bode's law; in fact, the Bode's law prediction for Neptune fits Pluto better! Therefore, either we must assign the law a more limited role than was initially hoped, or we search for a reason why it does not hold for the outermost planets, and for a modification to make it serviceable there also.

The latter alternative is the more fruitful of the two, but it must remain an unfulfilled wish until we can explain Bode's empirical, we might almost say numerological, law on some broader basis. Unless we know *why it holds* for Uranus, we cannot know *why it breaks down beyond Uranus*.

This may be a good place to review very briefly the different types of laws which we have encountered here. At the first level of sophistication there is the type of law which we have just called *empirical* and which seems to summarize simply some fairly directly observed regularity, e.g., Bode's law, or Kepler's laws.\* But as we have seen in our discussion of Kepler's work, even a regularity appearing rather obvious to us now may have been an insight which presupposed a difficult reorientation within science and which stimulated important further developments. At any rate, we must allow these rules a vital place in science.

Next in order, although any strict division is quite artificial, we may perhaps place those laws which represent an induction of one regulating principle from a variety of apparently very different phenomena. For example, Galileo's law of projectile motion and Newton's law  $F = ma$  were perceptions not obvious from direct observation. In fact, in each case the formulation of the law had to be preceded by the definition of new concepts, in Galileo's case, acceleration and the independence of horizontal and vertical velocity components, in Newton's case, mass and to some extent force. Because

\* Also Boyle's law of gases, Hooke's law of elasticity, Snell's law of refraction, Balmer's formula in spectroscopy, Ohm's law of electricity, Gay-Lussac's law of combining gases, the law of definite proportions in chemistry, and the laws of intrusion in geology.

these laws usually carry with them the definition of fundamentally important concepts involved therein, we might call them *definitional* laws (although that word has deceptive overtones). Examples of this type might include the conservation laws of energy and of momentum.

Thirdly, we turn to the remaining group, those laws which represent a general conclusion derived *from some postulate or theory*, whether new or old. For example, the lens laws which permit the design of optical instruments can be derived from a theory of the propagation of light. The law of pendulum motion can be deduced from the conceptual scheme of mechanics built around Newton's three laws of motion. And the law of universal gravitation certainly was not an empirical rule, nor did it serve to exhibit a new fundamental concept (all the factors in  $F = (Gm_1m_2)/R^2$  were previously defined concepts,  $G$  being essentially a proportionality constant); instead, as we have seen, it was a generalization derived from the conceptual scheme featuring Kepler's rules, Newton's own laws of motion, and the hypothesis of mass attraction among the planets and the sun. This type of law, which we may perhaps name *derivative* (sometimes called *causal*), often seems the most satisfying of the three because, since it has been derived from some underlying theory, we are tempted to feel it is also somehow more fully "explained." Scientists therefore continually seek to reduce purely empirical laws to derivative ones, thereby giving the empirical law a measure of additional "meaning,"\* while at the same time extending and fortifying the theory itself. There is a kind of symbiotic relationship here between law and theory. A theory becomes more and more respected and powerful the more phenomena can be derived from it, and the law describing these phenomena becomes more meaningful and useful if it can be made part of a theory. Thus, Newton's theory of universal gravitation gained greatly in stature because it enabled one to derive the laws governing the motion of the moon, which had been known by empirical rules since the days of the Babylonian observers. These rules in turn, now that they were "understandable" in terms of the theory, could be reformulated more accurately; in fact, the

---

\* This feeling appears somewhat fallacious if subjected to rigorous analysis. After all, the theory itself is acceptable only if its derived laws can be checked by experiments, and may in fact be incorrect *even if* the derived laws are correct (we shall see such an example when we come to deal with the theory of light). Nevertheless, great theories are relatively stable for long periods, and laws derived from them are as secure as we have any right to expect.

theory even enriched them, e.g., by calling attention to hitherto unnoticed peculiarities of the moon's motion. Another important example of this kind will be presented in the section on the nuclear atom; that theory has been so enormously important in modern science because it swallowed up whole textbooks full of isolated empirical and definitional laws from all branches of physics and chemistry, and reissued them as derivative laws springing from the unifying concept of the atom.

From this discussion we may perhaps draw some general conclusions valid beyond the physical sciences. Empirical rules may be valid enough in a limited field, but they do not contain enough information to warn the naive user when some particular new case is really outside the field of applicability of the empirical rule. Not until the damage is done is it clear that the rule did not apply.

Most of our own everyday actions tend to be governed by this type of uncomprehended rule and plausible generalization. This is doubly unfortunate as these rules, unlike those in science, are usually not even based on accurate clear-headed observation. While our life thereby may become simpler, it also is made more brutal and senseless, and we open ourselves up to an invasion of pseudoscience, superstition, and prejudice, even in matters subject to test and measurement. A concrete example may help to illustrate this point: A certain color of tie, eye, or skin may indeed have characterized some past acquaintances whom I or my friends did not accept as intellectual equals, but in the absence of impartial proof that native intelligence depends on those colors (and particularly in the face of proof to the contrary) I shall have to revise my attitude toward these men, and instead of a generalized judgment on all as one group I must now be prepared to make individual judgment on each; perhaps I shall even be moved to investigate why my previous experiences with this group revealed that so many men had not been given an adequate education to develop their abundant natural talents.

To return to Bode's empirical rule, recent work has given hope that its features will be explained by (derived from) Newton's law of celestial mechanics. As hinted before, the mutual perturbations of planets over billions of years would leave only a few orbits as stable possibilities; and some cosmologists and astrophysicists at present seem to be near the inevitable proof that the celestial laws predict such stability for the orbits that are actually observed. But this is a most difficult mathematical problem which from time to time has challenged some of the greatest scientists during the past 100 years.

PROBLEM 11-17. What can Bode's law predict concerning a hypothetical planet between the Sun and Mercury? (On the basis of a somewhat doubtful single observation in 1859 it was once thought to exist and was given the name *Vulcan*, but has never been found since.) If suddenly an additional small planet were seen between Jupiter and Saturn, might Bode's law be recast in some other form to hold for all planets? Might there be a form of Bode's law for the Ptolemaic system? Is Bode's law true? Is it false?

PROBLEM 11-18. Assume that in another planetary system than our own, the distance from the central star to each of the four inner planets is given by the following table:

Sequence number of planet ( $n$ ):	1	2	3	4
Radius of orbit, in $10^6$ miles ( $R$ ):	3	5	9	17

Obtain from these data a "law," a formula that will relate  $R$  to  $n$  for each planet. At what distance do you expect the fifth planet to be?

(g) Even in Newton's work we find further consequences of the law of universal gravitation. For example, he gave an accurate explanation for the long-standing mystery of the precession of the equinoxes, i.e., a very slow rotation of the earth's axis much like the wobbling motion of a rapidly spinning top. But now we turn to a more ambitious question. Do Newton's laws, so serviceable and fruitful within the solar system, continue to apply beyond, among the "fixed" stars?

To Copernicus or Galileo this question would have been meaningless, for it was not until Newton's time that relative motions were noticed in the celestial sphere (the discoverer was Halley). In fact, our whole solar system was found to move with respect to those distant stars. William Herschel, in 1803, also discovered that some star neighbors rotated about each other (double stars), and his son showed that their motions are compatible with the assumption of central forces between them, such as exist in the solar system.

This new picture of our living, moving universe, which the prophetic Giordano Bruno had sensed long ago, places the solar system in the uncrowded company of many billion other suns and their possible attendants (of which less than 6000 suns are visible without telescope). The whole cloud of stars forms our *galaxy*, a vaguely lentil-shaped region about 100,000 light years\* across and about 1500 light years thick. At a distance of about 30,000 light years from

\* One light year is the *distance* traveled by light in one year in vacuum, i.e.,  $9.46 \times 10^{12}$  km or  $5.88 \times 10^{12}$  miles.

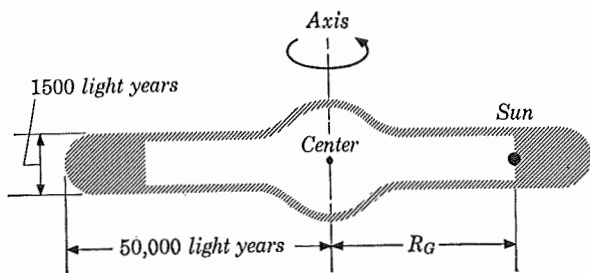


FIG. 11-7. Position of our sun in the galaxy.

the galactic center (not even at the center of all solar systems!) our planetary system is rather lost in this structure.

Such a picture evokes the thought that the mutual gravitational attraction of the stars should make the galaxy slowly gather up and coalesce into one solid mass at the center. Since this does not happen, motion of the stars being not all toward the center, there are then two explanations. Either the law of gravitation does not extend itself to the whole universe or, in analogy with the solar system itself, the whole galaxy is spinning around its axis, "distracting" the threatening gravitational attractions by making them do the chores of centripetal forces.

On examination, the second choice actually seems to hold.\* Our own solar system revolves about the distant center of the galaxy at a speed of about 150 miles/sec. According to our present view, the edge of the whole galaxy, whose framework we see as the Milky Way.† completes one revolution in perhaps a quarter-billion years.

Now let us make a bold little computation. If Newton's laws hold, we should be able to compute the approximate mass of the whole galaxy from the rate of (circular) revolution of our own sun about the galactic center.

If we neglect that part of our galaxy lying in the shaded portion in Fig. 11-7, we may say that the centripetal force on our sun is the gravitational pull exerted on it by the mass of the galaxy (acting approximately as if that were all located at the center). Thus we expect that the following equation holds:

\* For a modern view of the structure of our galaxy and the universe beyond see B. J. Bok and P. F. Bok, *The Milky Way*, Philadelphia, Blakiston Co. (1945); H. Shapley, *Galaxies*, Philadelphia, Blakiston Co. (1943); a short summary by Eddington is in Shapley, Wright, and Rapport, *Readings in Physical Sciences*, pp. 122-128, New York, Appleton (1948).

† Galaxy is derived from the Greek word *gala*, milk.

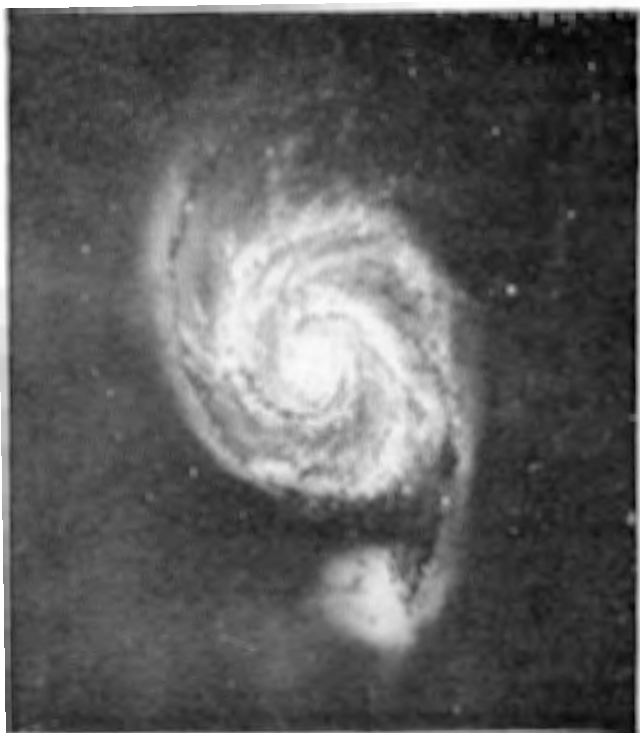


FIG. 11-8. Spiral nebula, a galaxy about a million light years from our own. (Photographed with the 200-inch telescope on Mt. Palomar.)

$$m_s \frac{4\pi^2 R_G}{T^2} = G \frac{m_s m_G}{R_G^2},$$

where  $T$  stands for the period of revolution of the sun (about  $2 \times 10^8$  years),  $m_G$  for the mass of the galaxy,  $m_s$  for the mass of the sun. The only unknown in this equation is  $m_G$ , and on solving for it,  $m_G = 2 \times 10^{11}$  times the mass of the sun (check this).

But when we add up the mass of all the stars which our galaxy is likely to hold, even on the most generous estimate, we can account for less than half of those 200 billion units. Either our assumptions are wrong, or we can find some galactic mass other than that contained in the stars of our galaxy. The assumptions were, of course, rather crude, but short of abandoning the law of gravitation, no tinkering with those admittedly qualitative details will explain the calculated overabundance of mass. So we turn to the alternative—and certainly we *can* find evidences that our galaxy contains matter not

congregated into suns and solar systems, namely, nebulae, interstellar dust, and most rarefied gases.

There are many such clues to the presence of this tenuous matter that would rescue the Newtonian laws, mainly its effect on starlight. But the density is inconceivably low, less than for the best vacuums in the laboratory, perhaps only a few atoms per  $\text{cm}^3$ . But this is spread over equally inconceivable distance, and an estimation of the total distributed material supplements the known star masses by about the missing amount. Thus our confidence in the law of gravitation rises to galactic proportions.

Beyond our world, starting at a distance of about thirty times our own galactic diameter, we find other galaxies scattered through space, as far as our biggest telescopes can reach. These island universes do not seem to differ greatly in size from ours and often appear like a spiral or pinwheel (see Fig. 11-8), probably the shape of our own galaxy also. (It is clear that if Kepler's third law holds for a whole galaxy, the outlying material will revolve more slowly than the masses near the center, hence the possibility of the spiral shape. On the other hand, if the whole system were to rotate like a solid disk, each part having the same period of revolution and angular velocity, then the centripetal force holding the system together could not obey an inverse square law but would have to be linearly proportional to the distance from the center.)

Even as these distant galaxies seem to recede from us and from one another at the speed of 100 to 1000 miles/sec, they rotate at the same time at rates close to our own, hence may be assumed to contain about the same mass and to obey the same laws of mechanics. And with this extension of the law of gravitation to the expanding Universe, "to all the distance of nature's infinitude," we find ourselves at the limit of our imagination, and at the height of admiration for so universal a principle. There have been few products of human genius to match its ambitious promise and its exciting fulfillment.

**11-7 "I frame no hypotheses."** The *theory* of gravitational forces, whose main hypothesis is the attraction of all particles of matter for one another, yields the derived *law* of universal gravitation, which in turn explains, as we have seen, Kepler's empirical laws and a wealth of other phenomena. Since one purpose of any theory is this type of explanation and summary, Newton's theory strikes us as eminently satisfactory. But there remained one feature which gravely bothered Newton's contemporaries and, indeed, students to our day.

How could one account for gravity itself? What is it that causes the attraction of one body for another? Is there not some intervening medium which somehow transmits the pull in a mechanical fashion?

The very statement of these questions reflects how firmly the mind is committed to physical models and to mechanical explanations, and how unsatisfied a mathematical, abstract argument leaves our emotions initially. Rather than accept "action at a distance," i.e., the idea that bodies can exert forces on one another without the agency of a medium between them, most scientists preferred to think of all space as filled with some kind of omnipresent fluid, which, apart from being otherwise so tenuous by itself as to be undetectable by any experiment, still had to be strong and versatile enough to communicate all gravitational forces (and incidentally perhaps also serve to propagate light and electric and magnetic effects).

The search for some such medium was long and the arguments loud. At the end of Book III of the *Principia*, Newton put his often misinterpreted remarks:

"But hitherto I have not been able to discover the cause of those properties of gravity from phenomena [observation and experimentation], and I frame no hypotheses . . . To us it is enough that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies and of our sea."

This is a famous statement, clearly an echo of Galileo's admonition in very similar circumstances (see Section 2-3). With the words, "I frame no hypotheses," Newton exempts himself from the obligation to account for the observed consequences of the gravitational theory by additional hypotheses (for example, of a mechanical ether) beyond those needed to derive the laws and observations. The mathematical law of gravity explains a wide range of observations; that is enough justification for accepting it. It is easily conceivable that one might eventually explain the law itself in terms of something still more fundamental, but that, Newton suggests, he is at the time unprepared to do. Nor does he feel that his theory suffers by this inability, for the purpose of scientific theory is not to find final causes and ultimate explanations, but to explain observables by observables and by mathematical argument; *that* he has done.\*

\* For a good discussion of Newton's much-debated statement, see F. Cajori's Appendix, Note 55, to his revision of the Motte 1729 translation of the *Principia*; also pp. 183-190, W. Dampier, *A History of Science*, New York: Macmillan (1936 edition).



On the other hand, his refusal, in the absence of experimental clues, to propose a mechanism by which the effect of gravity may pass from one body to another does not mean he will permit the opposite error, namely, dismissal of the question by the invention of some principle of gravity innate in matter, such as might well have satisfied a medieval scholar. This is brought out in a passage in the *Opticks*, in a discussion on the possible motion of the smallest particles of matter:

“. . . And the Aristotelians gave the Name of occult Qualities not to manifest Qualities, but to such Qualities only as they supposed to lie hid in Bodies, and to be unknown causes of manifest Effects: Such as would be the Causes of Gravity, and of magnetick and electrick Attractions, and of Fermentations, if we should suppose that the forces or Actions arose from Qualities unknown to us and incapable of being discovered and made manifest. Such occult Qualities put a stop to the Improvement of natural Philosophy, and therefore of late Years have been rejected. To tell us that every species of Thing is endow'd with an occult specifick Quality, by which it acts and produces manifest Effects, is to tell us nothing: But to derive two or three general Principles of Motion from Phaenomena, and afterwards to tell us how the Properties and Actions of all corporeal Things follow from these Principles would be a very great step in Philosophy, though the Causes of those Principles were not yet discovered: And therefore I scruple not to propose the Principles of Motion above-mentioned, they being of very general Extent, and leave their Causes to be found out.”

Nevertheless, the question of how one body can be acted on by another across empty space has an obstinate appeal for our picture-seeking mind. Part of the great following of Descartes' vortex theory, the scheme generally accepted during most of Newton's life, was that it provided a plausible picture of a universe completely filled with a whirlpool of material corpuscles whose action on one another and on the planets was simple physical contact.\* There we have an example of a plausible, easily grasped conceptual scheme which had the fault, fully appreciated by Newton but not by the majority of his contemporaries, of failing to agree quantitatively with observations, specifically, of failing to allow the derivation of Kepler's laws.

\* See H. Butterfield's *Origins of Modern Science* for a brief but stimulating account of the long struggle between Newton's theory and Descartes'.

But by contrast with the most successful feature of Descartes' picture, its inherent appeal to the common sense of the muscles, Newton's lack of an intuitive "reason" for planetary motion stood out that much more glaringly. Newton had to spend much time in his later life to uphold his less intuitively appealing system against the Cartesians, and at the same time to refute insistent attempts within his own camp to introduce metaphysical implications into the law of universal gravitation. He declared again and again that he was neither able nor willing "to adjudge natural causes" of gravitation. This was all the more remarkable (and a large step toward the modern conception of what is required of physical theory) since Newton, who, we must remember, later regarded himself as much theologian as scientist, might have introduced an explicit theistic assumption at that point. As has been hinted before, the existence of a Creator is a persistent implicit hypothesis in his work, and at times it does come out into the open, rarely in the *Principia* but more often in his other writings. Thus he could say in the *Opticks* in explanation of the term "absolute space," "God moves the bodies within his boundless uniform sensorium." Privately, Newton, like his contemporaries, did in fact believe that some material agency would eventually be found to explain the action at a distance.

There are several strong hints in his writings that he entertained the idea of an all-pervading ether, e.g., *Opticks*, III, query 18, where he asks whether heat is not "conveyed through the *vacuum* by the vibration of a much subtler medium than air, which, after the air was drawn out [as by evacuating a vessel with a pump], remained in the *vacuum*"; this medium he imagined as "expanded through the heavens." At the time, this was a most reasonable speculation, proceeding as it did from some simple observations on the transmission of (radiant) heat to a thermometer suspended in an evacuated vessel. It is the more interesting that the tempting but vague picture was not permitted in the *Principia*.

**11-8 Newton's place in modern science.** So impressive were the victories of Newtonian mechanics in the early eighteenth century that, through the customary extrapolation from science to philosophy, there spread a mechanistic world view which asserted that man's confident intellect could eventually reduce *all* phenomena and problems to the level of mechanical interpretation. In economics, philosophy, religion, the "science of man"—everywhere the success of

the work of Newton and the Newtonians was one of the strongest influences to encourage the rising "Age of Reason." \*

One of the consequences of the mechanistic attitude, lingering on to the present day, was a widespread belief that with the knowledge of Newton's laws (and those of electrodynamics later) one could predict the whole future of the whole universe and each of its parts if only one were given the several positions, velocities, and accelerations of all particles at any one instant. It was a veiled way of saying that everything worth knowing was understandable in terms of physics, and that all of physics was essentially known. Today we honor Newtonian mechanics for less all-inclusive but more valid reasons. The factual content of the *Principia* historically formed the basis of most of our physics and technology, and the success of Newton's approach to his problems revealed the most fruitful method to guide the work of the subsequent two centuries in the physical sciences.

We also recognize now that despite its breathtaking scope of applicability, Newton's mechanics holds in its stated form only within a definable region of our science. For example, although the forces *within* each galaxy may be Newtonian, one can speculate that forces of repulsion instead of attraction operate between each galaxy and its neighbors (one interpretation of the greater speed with which more distant systems recede). And at the other end of the scale, among atoms and subatomic particles, an entirely non-Newtonian set of concepts has to be presented to account for the behavior of those small-scale worlds.

Even within the solar system, there are a few small residual discrepancies between predictions and facts. The most famous is the erroneous value of the position of closest approach to the sun by the planet Mercury; calculations and observations differ by some 43 seconds of arc per 100 years. A similar small failure of Newton's laws of motion and of gravitation exists for the path of a recurring comet that comes close to the sun and, to a progressively lesser degree, for the rest of the members of the solar system.

But these difficulties cannot be traced to a small inaccuracy of the law of gravitation (e.g., to a slight uncertainty of the exponent in the expression  $Gm_1m_2/R^2$ ). On the contrary, as in the case of the failure of the Copernican system to account accurately for the detail

---

\* A most stimulating discussion of this important aspect of science will be found in J. H. Randall's *The Making of the Modern Mind*, Chapters X-XV, Boston: Houghton Mifflin Co. (1940).

and "fine structure" of planetary motion, we are here again confronted with the necessity of scrutinizing our assumptions. Out of this study has come the realization that no single revision can account for all these deviations from classical predictions, so that at present Newtonian science is joined at one end with relativity theory which, among many other achievements, covers phenomena involving bodies moving at high speeds and/or the close approach of large masses. At the other end Newtonian science borders on quantum mechanics, which gives us a physics of atoms and molecules. As for the middle ground, Newtonian mechanics still describes the world of ordinary experience and of "classical" physics as accurately and satisfyingly as it did always.

### Additional Problems

PROBLEM 11-19. Would Newton's laws of motion, as usually stated, hold in a geocentric system like Ptolemy's? (*Hint:* The term "motion along a straight line" in Newton's first law refers to a line in space such as might be drawn from the sun to a relatively "fixed" star.) If they do not, explain why not and whether it is plausible that equivalent laws might be formulated to describe motions in a geocentric system.

PROBLEM 11-20. Turn to Section 8-2 and use those paragraphs to rate the Newtonian theory of universal gravitation point by point. (Don't assume that such a conscious rating is ever seriously undertaken in scientific work.)

PROBLEM 11-21. Turn to page 112, where there are in the first paragraphs 5 objectives for all of Part C. Examine each objective and illustrate how it was developed.

### Further Reading

E. A. Burt, *The Metaphysical Foundations of Modern Physical Science*. London: Kegan Paul (1932). Invaluable. See especially Chapters III and VII.

H. Butterfield, *op. cit.* (Ch. 2), Chapter VIII.

Sir Oliver Lodge, *Pioneers of Science*. London: Macmillan (1926, many previous editions). See specifically Lectures 6 to 17.

W. F. Magie, *op. cit.* (Ch. 2), pp. 92 and 105-111. Excerpts from Newton and Cavendish. Also W. S. Kniekerbocker, *op. cit.*, Chapters IX (Newton) and XV (Herschel); J. W. Knedler, *op. cit.*, pp. 198-243 (*Principia*); H. Shapley and H. E. Howarth, *op. cit.*, pp. 74-93 (*Principia*), 140-151 (Herschel), 180-182 (Bode), 245-248 (Adams), 249-254 (Leverrier); and H. Boynton, *op. cit.*, pp. 104-117 (Cavendish and Herschel).

Isaac Newton, *op. cit.* (Ch. 4), particularly the beginning of Book Three and the General Scholium, including the notes.

*Sir Isaac Newton, 1727-1927.* Baltimore: Williams and Wilkins (1928). Essays on different aspects of Newton's work.

F. A. Saunders, *A Survey of Physics for College Students.* New York: Henry Holt and Co. (1943). An authoritative text; see Chapter 8 for experimental measurement of  $G$ .

D. Stimson, *Scientists and Amateurs.* New York: Henry Schuman (1948). A penetrating account of the Royal Society and its illustrious members.

F. S. Taylor, *op. cit.* (Ch. 6). See Chapters V through VIII, and XV.

F. G. Watson, *Between the Planets.* Philadelphia: Blakiston Co. (1941).

F. L. Whipple, *Earth, Moon and Planets.* Philadelphia: Blakiston Co. (1941).

A. Wolf, *A History of Science, Technology and Philosophy in the XVIIth and XVIIIth Centuries*, and same for *XVIIIth Century.* New York: Macmillan (1935 and 1939). Generally useful for its historical material and illustrations.

## *Part D*

# ON STRUCTURE AND METHOD IN PHYSICAL SCIENCE

- 12 Concepts
- 13 The Growth of Science
- 14 Laws

“As in Mathematicks, so in Natural Philosophy, the Investigation of difficult Things by the Method of Analysis, ought ever to precede the Method of Composition. This Analysis consists in making Experiments and Observations, and in drawing general Conclusions from them by Induction, and admitting of no Objections against the Conclusions, but such as are taken from Experiments, or other certain Truths. For Hypotheses are not to be regarded in experimental Philosophy. And although the arguing from Experiments and Observations by Induction be no Demonstration of general Conclusions; yet it is the best way of arguing which the Nature of Things admits of, and may be looked upon as so much the stronger, by how much the Induction is more general. And if no Exception occur from Phaenomena, the Conclusion may be pronounced generally. But if at any time afterwards any Exception shall occur from Experiments, it may then begin to be pronounced with such Exceptions as occur. By this way of Analysis we may proceed from Compounds to Ingredients, and from Motions to the Forces producing them; and in general, from Effects to their Causes, and from particular Causes to more general ones, till the Argument end in the most general. This is the Method of Analysis: And the Synthesis consists in assuming the Causes discover’d, and establish’d as Principles, and by them explaining the Phaenomena proceeding from them, and proving the Explanations . . .”

Isaac Newton, *Opticks*.

## ON STRUCTURE AND METHOD IN PHYSICAL SCIENCE

As long as there has been science, there has also been commentary on the tasks and procedures of science. From Aristotle to Galileo, from Roger Bacon to the present, almost all major scientists and philosophers have contributed their varying opinions. No room is left, it would seem, for remarks on a subject so lengthily and so competently examined. But even if this were true, we might permit ourselves to discuss, in an informal manner, representative opinions held by some of the men who have prominently influenced scientific thought, and so attempt to reach a plausible point of view from which to analyze the progress and procedures of science—not for the sake of a technical exposition of the problems involved, but for increased enjoyment and understanding of the scientific enterprise.

The three chapters in Part D carry on and complete the cycle begun with Chapter 8, *The Nature of Scientific Theory*.

## CHAPTER 12

### CONCEPTS\*

**12-1 Introduction: the search for constancies in change.** When you ask "What is science?", you are asking mainly "What do scientists now do at their desks and in their laboratories, and what part of their past work is still useful to men in a given field?" Because the physical scientist has, over the last four centuries, removed himself more and more from direct communication with nonscientists, let us first visualize his tasks in terms of an analogy involving his youthful and more immediately approachable colleague, the cultural anthropologist. In that case the laboratory, says Margaret Mead's *Samoa*, is an island of primitive people. The anthropologist immerses himself in the life of the community to observe it, to study its pattern—but, of course, the eye cannot see, the mind cannot grasp, meaningful patterns within the chaos of movement and sound until one has crystallized out of one's experience the necessary concepts with which to discover, think about, and describe relationships in the community life. One set of concepts such as *family*, *ruler*, *in-group*, *I.Q.*, etc., belongs to the prior "common-sense" background of our investigator—this corresponds in the physical sciences to the time-honored notions of *time*, *volume*, *speed*, and so on. Another and to us more interesting set of terms comes out of the phenomenon under study. The anthropologist will learn the language and customs of this people and will discover and invent new constructs important to the organization of his observations and his understanding of the community, e.g., the terms "taupo" and "uso," which denote in Samoa respectively the ceremonial princess of the house and a sibling of the same sex. In the same manner the physical scientist, with evidently greater difficulty, must extract from his experience new guiding ideas like *electric field* or *valence*.

Significantly, our anthropologist may find absent the manifestation of many concepts our own society seemingly could not dispense with, e.g., some elements of the moral code of the West, the knowledge of one's own exact age, or even the simple recognition and naming of most colors; and again science, too, had to learn—often the hard way—that there are no useful equivalents in inanimate nature for such common-sense terms as human longing or absolute simultaneity.

---

\* It is intended that Chapters 12, 13, and 14 be re-read on reaching the end of the book.



The formulation or "discovery" of new concepts helps the anthropologist gradually to reinterpret the originally meaningless, aimless, or "primitive" life of the village, and a complex, rather rigid pattern emerges. His own early common-sense concepts themselves may undergo profound changes—the word *law*, for example, is surrounded by an entirely different atmosphere in different cultures, just as the idea of *force* in physics (and, in fact, every technical concept that grew out of the vocabulary of common sense) is now at variance with the original meaning of the same word as still used outside the sciences.

In the end, our anthropologist may succeed in decoding the original problem, bringing back an account of that people's political and family organization, personal esthetic values, religious beliefs and practices, economic methods, and so on. Perhaps he will also be able to reconstruct the history of that people, or even to illuminate the behavior of another group, his own, in relation to that culture. At any rate, it is evident in this example that his job was not finished when he had collected the initial direct observables; on the contrary, that was the barest beginning, the stimulus for the really important part of his work. We may keep in mind, with all its limitations, the analogous picture of physical scientists as explorers in a universe of events and phenomena, attempting to find its pattern and meaning.

**12-2 Science and nonscience.** Science, one may now say, is the ever unfinished quest to discover all facts, the relationships between things, and the laws by which the world runs. But let us go beyond this. While not trying to propose a one-sentence definition of the whole complex concept "science," we may perhaps agree at the outset that the main business of science is to trace in the chaos and flux of phenomena a consistent structure with order and meaning, that is, to interpret and to transcend direct experience. "The object of all sciences," in Einstein's words, "is to coordinate our experiences and to bring them into a logical system." And Niels Bohr agrees when he says "The task of science is both to extend the range of our experience and to reduce it to order."

Probably you find these statements to be too all-inclusive; the same aim might well be claimed by art or by philosophy. Thus T. S. Eliot has recently said, "It is the function of all art to give us some perception of an order in life by imposing an order upon it," and A. N. Whitehead defined speculative philosophy as "the endeavor to frame a coherent, logical, necessary system of general ideas in terms

of which every element of our experience can be interpreted." Indeed, in science, as in art and philosophy, our most persistent intellectual efforts are directed toward the discovery of pattern, order, system, structure, whether it be as primitive as the discernment of recurring seasons or as sweeping as a cosmological synthesis. The search for constancies in the flux of experience is so fundamental and so universal a preoccupation of intelligent life itself that, in common with many of the Greek philosophers, we may regard mind as the principle that produces order. We note without astonishment among the great early men of science an artist and a churchman, Leonardo da Vinci and Copernicus. Indeed, the very origin of the word *science* (Latin *scire*, to know, to learn) betrays the extent of its appeal.

Of course, one must not deny the fundamental and distinct differences separating the sciences from the nonsciences. There are obvious points which set one apart from the other: to predict nature and so to control her is the motivation for comprehension of nature through science, whereas the poetic understanding of nature is self-realization, a proclamation and ennoblement of man's own spirit. But these are in a sense complementary, not contradictory, motivations, illuminating the two sides of man. Both co-exist to some degree within each individual. How artificial would be a distinction between the triumph of the scientist and that of the artist within Johannes Kepler as he writes in the *Harmony of the World* (1618):

" . . . What I prophesied 22 years ago, as soon as I found the heavenly orbits were of the same number as the five [regular] solids, what I fully believed long before I had seen Ptolemy's Harmonies, what I promised my friends in the name of this book, which I christened before I was 16 years old, I urged as an end to be sought, that for which I joined Tycho Brahe, for which I settled at Prague, for which I have spent most of my life at astronomical calculations—at last I have brought to light, and seen to be true beyond my fondest hopes. It is not 18 months since I saw the first ray of light, 3 months since the unclouded sun-glorious sight—burst upon me. Let nothing confine me: I will indulge my sacred ecstasy. I will triumph over mankind by the honest confession that I have stolen the golden vases of the Egyptians to raise a tabernacle for my God far away from the lands of Egypt. If you forgive me, I rejoice; if you are angry, I cannot help it. This book is written; the die is cast. Let it be read now or by posterity, I care not which . . ."

A second point of difference between science and nonscience is much more clear-cut than the first; it is the kind of rules and concepts the scientist uses, the type of argument which will cause him to say, "Yes, I understand and I agree." This will occupy our attention to some degree, as will a third point of difference: the observation that in the course of time, despite great innovation and revolutions, there accumulates in science a set of internationally acceptable, basic, and fairly enduring conceptual schemes, whereas this can hardly be said for many other human endeavors.

At once the question arises as to how man, so short-lived and so fallible, can penetrate such complex truths, discover lasting patterns, and obtain general agreement. For the scientist's work, like that of any other explorer, must involve the whole man, demanding as it does reflection, observation, experimentation, imagination, and a measure of intuition. Being human, he fails far more often than he succeeds, and even his successes and failures may in the light of further progress reverse positions. He often cannot fully explain his reasons for dedicating himself to his work and the steps by which he progresses; and if he is vocal on such matters, he very likely is contradicted by the testimony of some of his colleagues. And yet, the *result* of this uncertain human activity, namely, the growing body of science itself, is undeniably a successful, vigorous enterprise, beautiful in discoveries, in distinct contrast to the frailties and confusions of its human creators. This paradox we shall explore, for it will be the key to an understanding of the activities of the scientist and of the successful features of his work.

Immediately an answer suggests itself: there must be—so goes the reasoning of some scholars and laymen—some trick, some *method* which scientists follow to solve their problems so well. After briefly disposing of this opinion, we shall look quite closely at an alternative: Is there perhaps some special merit in the idiosyncratic vocabulary, in the particular mental tools themselves with which scientists analyze the world around them?

**12-3 The lack of a single method.** If by scientific method we mean the sequence and rule by which scientists now and in the past have actually done their work, then two truths soon become obvious. First, as for every task, there are here not one but many methods and uncountable variants and, second, even those different methods are for the most part read into the story after it has been completed, and so exist only in a rather artificial and debatable way. The ever-

present longing to discover some *one* master procedure underlying all scientific work is understandable, for such a discovery might enormously benefit all fields of scholarship; but like the search for the philosopher's stone, this hope had to be given up. At about the beginning of the 17th century, the time of Francis Bacon and Descartes, it was still quite reasonable to hope that both these all-powerful keys to wealth, health, and wisdom existed, but the verdict of the three centuries since then is plainly negative. Even 175 years ago the chemist Joseph Priestley warned ". . . how little mystery there really is in the business of experimental philosophy, and with how little *sagacity*, or even *design*, discoveries (which some persons are pleased to consider as great and wonderful things) have been made . . ."\*

Priestley's words, even though an extreme generalization which does not do justice to his own carefully prepared researches, are a proper antidote to the other extreme of opinion which would present science to us as a special, infallible scheme relentlessly progressing from success to success with the precision of a smoothly moving machine. For as soon as one begins to look into the history of scientific discoveries, one is overwhelmed by evidences that there is no single, regular procedure. By temperament and by characteristics of performance, scientists have always differed from one another as widely as composers. Some proceeded from one step to the next with the certainty and restraint of a Bach, others moved among ideas with the abandon of a Schumann. Among the great scientists there have been adventurers and recluses, self-taught mechanics and aristocrats, saints and villains, mystics and business men, pious men and rebels. In *The Study of the History of Science*, George Sarton said about them:

". . . Their manners and customs, their temperamental reactions, differ exceedingly and introduce infinite caprice and fantasy into the development of science. The logician may frown but the humanist chuckles.

"Happily such differences are more favorable to the progress of science than unfavorable. Even as all kinds of men are needed to build up a pleasant or an unpleasant community, even so we need all kinds of scientists to develop science in every possible direction. Some are very sharp and narrow-minded, others broad-minded and superficial. Many scientists, like Hannibal, know

---

\* From his Volume II of *Experiments and Observations on Different Kinds of Air* (1776), quoted in Case 2 of the *Harvard Case Histories in Experimental Science*, Harvard University Press (1950).

how to conquer, but not how to use their victories. Others are colonizers rather than explorers. Others are pedagogues. Others want to measure everything more accurately than it was measured before. This may lead them to the making of fundamental discoveries, or they may fail, and be looked upon as insufferable pedants. This list might be lengthened endlessly."

The process of discovery itself has been as varied. While it is perfectly true that individual research projects are almost always unspectacular, each investigation being fairly routine, sound within its prescribed scope, and logically quite consistent, occasionally some most important theories have come from drawing wrong conclusions from erroneous hypotheses, or from the misinterpretation of a bad experiment. Sometimes a simple experiment has yielded unexpected riches, and sometimes the most elaborately planned experiment has missed the essential effect by a small margin. Even the historic conceptual schemes themselves first captured the mind in *seemingly* the most casual or unpredictable way. To cite some famous examples, Darwin said he got "a theory by which to work" while he happened to read Malthus' treatise on population "for amusement." One of Kekulé's fundamental contributions to chemistry came to him suddenly during an idle daydream. Similar things have happened to scientists in every field.

Moreover, some men had all the "significant facts" for an important finding in their hands and drew conclusions which were trivial or proved to be mistaken, whereas sometimes others established correct theories although some facts before their very eyes were seemingly in violent contradiction. Of all these we shall have further examples. Even the work of the mighty Galileo, viewed in retrospect, sometimes seems to jump ahead from error to error until the right answer is reached, as if with the instinctive certainty of a somnambulist.

In Chapter 14 we shall hear the opinions held by the men of science themselves concerning the methods of work in their field, but even this superficial recital of variety is enough to establish a feeling that the success of science lies not so much in some single method of work, but perhaps rather in some peculiar adjustment or mechanism mediating between two factors—the achievement of the individuals in a field, and the body of science in which their contributions are brought together after supplementing and modifying, cooperating and competing with one another. A very crude analogy may help here; if we compare the structure of science to that of an ant hill,

the successful construction of that astonishingly complex, well-arranged, and successful building surely must primarily be explained not by the diverse, almost erratic behavior of the individual insects, but rather by the marvelous coordination of innumerable solitary efforts. The mechanism of coordination among these social insects is still largely a mystery, but in the analogous case of science something positive might be said. We approach it by turning to the alternative possibility posed at the end of the previous section, to an examination of the astonishing and many-sided power that resides in the concepts scientists use.

**12-4 Physical concepts: measurement and definition.** All intelligent human endeavor stands with one foot on observation and the other on contemplation. But scientists have gradually come to limit themselves to certain *types* of observations and thought processes. One distinctly striking "limitation" is the tacit desire among scientists to assure that they are in fact discussing the same concepts in a given argument. This desire for clarity is, of course, general, but here we find a particularly successful scheme of meeting it, even at the cost of some sacrifices.

To take a very simplified example: If the scientific problem were to find the length or volume of a cubical block, its electric charge or chemical composition, the specialists called in to solve this problem would all appear to understand clearly what is wanted because they all would independently go through quite similar manual and mathematical operations to arrive at their answers, even though these answers may not completely coincide. (If the word *similar* seems disturbing in the previous sentence, we may go so far as to claim that these men will, on mutual consultation, very probably all agree on *exactly* the same operations to follow.) As a consequence, their findings for, say, the length of one edge of a block may read 5.01 cm, 5 cm, 4.99 cm, and 5.1 cm. The impressive thing here is not that they seem to disagree on the exact length or on the accuracy, but that they do agree on the type of answers to give. They do not, for example, say "as long as a sparrow in summer," "five times the width of my finger," "as long as it is wide," and "pretty short." Evidently the words "length of one edge of a block" mean the same things to all four experimenters, and we may be fairly sure that on consultation sooner or later all four would convince themselves of one single result, possibly  $5.04 \pm 0.04$  cm. It should not matter which meter stick they happen to pick up, for all manufacturers of measuring equipment are under legal obligation to check their products against

carefully maintained universal standards. On the whole, almost every measurement made in the laboratory is in a sense such a comparison between the observables and some well-known standard accepted by the whole fraternity. Much of the success and rapid growth of science depends on these simple truths, for it is clear that in such circumstances the energies of investigators are not regularly drained off by fruitless arguments about definitions and rules of procedure; on the contrary, the labors of many men over centuries can combine in one advancing stream.

There remain, of course, large areas of possible disagreement among scientists, but such arguments can be settled, often by having recourse to some one series of measurements which both disputants (rightly or wrongly) acknowledge at the time to be decisive one way or the other. One of the impressive features of modern physical science is the rapidity with which most major differences of opinion in the field usually disappear.\* The "secret" of this successful harmony and continuity in physical science (which we idealized a little in the previous paragraph) to a large degree lies in the *nature of concepts and their definitions*. For example, the concept "length of an object" as *used* in science is ultimately defined by the very *operations* involved in making the measurement. The problem "What is the length of a block?" is for all practical purposes identical with the question "What is the difference between those two numbers printed on a specific meter stick which stand directly below two corresponding scratches, entered there to signify local coincidence with adjacent corners of the block?"

This last sentence contained an abbreviated example of what we shall call an *operational definition*, that of length, and although it seems ridiculously involved, and may indeed never be used explicitly, our four experts can regard it as the "true meaning" of the length of a block, available for examination if any dispute should arise. Ideally, each of the concepts used in the physical sciences can be made clear in terms of some such definition, and that is the mechanism whereby mutual understanding among scientists is made possible. For it is clearly more difficult to misinterpret action than words.

---

\* A comparison with the state of affairs in other disciplines may be unfair, but this is a good place to interject that the writings on the history and the philosophy of science themselves, in this book as in others, are traditionally ready subject matter for disagreement and debate, not only for a few years but perhaps forever.

At this point you are likely to have several disturbing questions. If you were tempted to object that the operational definition given here is far from the common-sense meaning of length, you would be quite right and would have hit on a very important truth: everyday notions seem clear and scientific terms mysterious. However, a little thought shows the opposite to be the case. The words of daily life are usually so flexible and undefined, so open to emotional color and misunderstanding, that our first task here is to get used to the specific vocabulary of the sciences and to the apparently picayune insistence on its *rigorous* use—that enormously successful habit which scientists borrowed from scholastic logicians. Then again, you are bound to have a vaguely uncomfortable feeling that such a definition only shows us how to measure according to some man-made convention, and does not tell us what length “really is”; once more you are perfectly correct, and once more we must try to make our peace with the limitations of modern science. Here again we are touching on a problem that is at the very core of science, namely, what reality means to a man in a laboratory. Surely we shall have to explore this very thoroughly. For the moment let us take the answer of that great French mathematician and philosopher of science, Henri Poincaré, who half a century ago illustrated the operational attitude towards physical concepts in this manner: “When we say force is the cause of motion we talk metaphysics, and this definition, if we were content with it, would be absolutely sterile. For a definition to be of any use, it must teach us to *measure* force; moreover, that suffices; it is not at all necessary that it teach what force is *in itself* nor whether it is the cause or the effect of motion.”

Lastly, you may be puzzled that a simple measurable like length should not be given with perfect exactness, that we should have said “the edge of this block is  $5.04 \pm 0.04$  cm long.” The term  $\pm 0.04$  cm is called the Probable Error in the measurement; that is to say, the chances are fifty-fifty that the next measurement of the same edge will read between 5.00 cm and 5.08 cm (see Section 1-1). Evidently all measurements, apart from simple counting, will contain some error. But although this uncertainty may be slight, how can we dare build an exact science on concepts defined by necessarily uncertain measurements? The paradox is resolved by two recognitions. First, this word “error” does not have in science the connotations *wrong*, *mistaken*, or *sinful* which exist in everyday speech. In fact, a measured value can be regarded as “exact” only if we know also just what range of values to expect on repeating the measure-



ment. Secondly—and we must repeat this frequently—science is not after absolute certainties, but after relationships among observables. We realize that the observables can be defined and measured only with some uncertainty; and we shall not demand more than that of the relationships between them, thus leaving the search for absolute truths to other fields of thought.

The consequence of even such a trivial-sounding operational definition as that of length may be quite startling. In the special theory of relativity, rigorous definitions of this type directed the course of thought to unexpected results, one being that the measured length of an object depends on how fast the object moves with respect to the observer, a finding that accounted for some most perplexing previous observations. As a result of the impact of Einstein's work, scientists have become aware of the operational nature of the useful concepts; it was then discovered that some tacitly accepted ideas were leading to serious contradictions in parts of physical theory because, by their very formulation, they could not be connected with any possible activities or operations in the laboratory. For example, the concepts of time and space were defined previously not by specific and exclusive reference to manipulations with meter sticks, light signals, pendulum clocks, and the like, but in some absolute, intuitive sense. The classic reference here is Newton's statement in the first pages of the *Principia*:

“Absolute, true, and mathematical time, of itself, and from its own nature, flows equably and without regard to anything external, and by another name is called duration.”

Note the phrase “without regard to anything external,” that is, without necessary relation to the rotation of a second hand on a real watch. By this very definition of “true” time, we could not hope to measure it. Or again,

“Absolute space, in its own nature, without regard to anything external, remains always similar and immovable.”

From the point of view of modern practical scientists, such statements, inherently without operational meaning, are sometimes called “meaningless,” perhaps a rather abrupt term but accurate in this limited sense. You may well ask whether Newton's science was not invalidated by such meaningless basic postulates, and the answer is no, simply because Newton did not, in fact, depend on the explicit use of these concepts in his own scientific work. (His reasons for

stating them at all were complex and can be traced to his philosophical motivation.)

Galileo, while much nearer to that science of the ancients which he helped to displace, saw quite clearly that science should be based on concepts that have meaning in terms of possible observations. Recall again his preface to the definition of acceleration: "We however have decided to consider the phenomena of freely falling bodies with an acceleration such as actually occurs in nature, and to make our definition of accelerated motion exhibit the essential features of this type of natural accelerated motion."

If we define as "meaningless" any concept not definable by operations, it follows that not only concepts but whole statements and even intelligent-sounding, deeply disturbing questions may turn out to be meaningless *from the standpoint of the physical scientist*. Here are a few examples: "What are the physical laws which would hold in a universe completely devoid of any material bodies?" "Which is *really* at rest, the sun or the earth?" "Will this table cease to exist while it is not being observed?" "What is length in itself, apart from measurements?" "Are there natural laws that man can never hope to discover?" and even "Is light made of corpuscles only or waves only?"

There are many more such questions listed by the American physicist and philosopher, P. W. Bridgman,\* who speculatively ventures to extrapolate this "operational view" beyond science by saying "I believe that many of the questions asked about social and philosophical subjects will be found meaningless when examined from the point of view of operations. It would doubtless conduce greatly to clarity of thought if the operational mode of thinking were adopted in all fields of inquiry . . ."

As has been implied throughout, not all concepts in science can be defined by *manual* operations. Instantaneous velocity, for example, was defined by the slope of a straight line, tangent to the distance *vs.* time graph; this is a mental or mathematical operation. But each such definition becomes a fairly unambiguous set of directives to all scientists who have worked in the field, and this explains why their vocabulary is not regularly the subject of dispute.

You will find it revealing—more than that, almost indispensable

---

\* In his *Logic of Modern Physics*, New York: Macmillan (1927). This short and influential book outlines a modern approach to science; in particular, the operational definition of concepts is discussed fully. The first 100 pages are excellent reading for students at this stage of study.

to a true understanding of physical concepts—to make for yourself operational definitions of all important terms, such as we have made in previous chapters for acceleration, force, mass, and so on. In this task also it will be wise to keep in mind Bridgman's simple statement: "The true meaning of a term is to be found by observing what a man does with it, not what he says about it."

**12-5 Primary and secondary quantities.** If it be true that science is based on operationally meaningful concepts, then there must be a large range of experience that is being neglected by science. Perhaps you begin to compare a scientist with a man who looks at the night sky by means of a very powerful telescope and who is thus enabled to examine a small region with extreme penetration while, however, sacrificing the chance to watch a whole world at one glance. In addition, we are apt to notice that with his telescope our scientist seems to watch only for the number of stars, their relative brightness, and the like, instead of scanning the skies for the sheer experience drawn from the beauty of the grand display that so impresses the casual observer.\*

The comparison is apt. Galileo drew the distinction between those experiences and concepts which might safely serve as the foundation stones of science and those which, having a measure of more *subjective* meaning, should from the point of view of science be regarded as sources of illusion and debate. He called them respectively *Primary and Secondary quantities*. The Primary quantities were direct observables that could be mathematically symbolized and measured, such as position and motion, the very elements that at the time had simple operational significance, to use the modern phrase. As Secondary were regarded the quantities then not accessible to instrumental measurement, those which were largely qualitative. By and large, the distinction, which impressed and was accepted by such followers of Galileo as Newton himself, still holds in the physical sciences. We might say that Galileo's decision reduced the extent of *eligible* experience to a small fraction of the scientist's total experience, but to precisely that fraction which he could quantify and therefore share fairly unambiguously with his fellows. The ideas of science

---

\* The analogy is suggestive in another way: the prospect of inevitable, successive refinements of his telescope reminds the researcher that his findings are always tentative, whereas the unaided (nonscientific) observer can easily persuade himself that his observations are final and can yield final truths.

look so stylized and unreal just because we demand of them that they help us describe those features of experience which the common-sense view of reality cares least about—measurement, mathematical manipulation, numerical prediction, and clear communicability—while failing to describe exactly those most prominent uses of everyday expressions, namely, our feelings, reactions, and other personal involvements.

Recent developments in branches of psychology and the social sciences are reminiscent of that early search for quantification in physics and chemistry. Throughout Europe, men of the early 17th century were saying that the qualitative ideas inherited from the ancients would have to give way to a new, quantitative way of describing nature. A case in point is Aphorism XCVIII of Francis Bacon's *Novum Organum* of 1620, where we find this complaint concerning contemporary thinking: "Now for grounds of experience—since to experience we must come—we have as yet had either none or very weak ones; no search has been made to collect a store of particular observations sufficient either in number, or in kind, or in certainty to inform the understanding, or in any way adequate . . . Nothing duly investigated, nothing verified, nothing counted, weighed, or measured, is to be found in natural history; and what in observation is loose and vague, is in information deceptive and treacherous."

**12-6 Mathematical law and abstraction.** The insistence on quantitative concepts must, of course, appear incomprehensible until we recognize that the work of the physical scientist is based on a faith as ancient as it is astonishing, namely, that nature works according to mathematical laws and that the observations are *explained* when we find the mathematical law relating the observations. For example, nothing more is left to be said about simple freely falling bodies once we have found that they move according to the law  $s = v_0t + \frac{1}{2}gt^2$ . Galileo expressed it this way: "Philosophy [we should call it science now] is written in that great book which ever lies before our eyes—I mean the universe—but we cannot understand it if we do not learn the language and grasp the symbols in which it is written. This book is written in the mathematical language, and the symbols are triangles, circles, and other geometrical figures [we should now add all other mathematical devices] without whose help it is impossible to comprehend a single word of it, without which one wanders in vain through a dark labyrinth." To Galileo, as to his contemporaries (e.g., Kepler and, to some extent, even Harvey) and to modern physical scientists, mathematical methods provide the

technique *par excellence* for ordering and comprehending nature. To logic, which for the scholastics was the main tool of investigation, is now primarily delegated the task of establishing the consistency of the scheme of mathematical demonstration, hypotheses, and observations. While Galileo and Kepler still felt compelled to announce this faith loudly and often, by now, through the unforeseeable fruitfulness of this view, it has become a fundamental, unexamined attitude among physical scientists.

By saying that laws are mathematical relations between observables, we mean, of course, that the general descriptions of physics are usually stated in the form "variable  $X$  is related to variables  $Y, Z$ , etc. by such and such a mathematical function." (Consider again, for example, the equation for free fall or Newton's law of gravitation.) Laws may state that  $(X/Y) + Z$  (or some other combination of the variables) always has a certain numerical value, or tends to attain a maximum or a minimum value under given conditions. Yet another example is the superposition principle, which asserts, you recall, that the two variables  $v_x$  and  $v_y$  in projectile motion are *not* interdependent. Above all, there is that type of law most eagerly sought which says in effect "this function of these variables under given conditions is always *constant*." The law of freely falling bodies as conceived by Galileo was of this kind, for it could be phrased:  $s/t^2$  is constant if the falling body starts from rest. Kepler's third law is an obvious example. There will be numerous others: the laws of conservation of mass, of momentum, of energy, of "caloric," etc. To us it is clear why laws of constancy are so highly prized; they combine the most successful features of science with its most persistent preoccupation—the mathematical formulation of concepts ( $s, t$ , etc.) that aid in the discovery of unchanging patterns in the chaos of experience.

The mathematical formulation of physical law has an incidental and initially disconcerting effect. Unless we are warned to think of a law as the relationship between variables, we are apt to regard it as a strict relationship between *cause and effect*. But in our law of the type  $X = YZ$  we can just as well write  $Y = X/Z$ , or  $Z = X/Y$ , and there is no way of telling whether  $X$  or  $Y$  or  $Z$  is a cause or an effect. As a consequence, it is on the whole more fruitful to think of an interaction rather than a simple causation, and to ask "to what factors is  $X$  related" instead of "what causes  $X$ ." For example, Boyle's law states that for a given mass of gas, at a constant temperature, the absolute pressure  $P$  and the volume  $V$  of the gas are

related by  $PV = \text{constant}$ , i.e.,  $P$  is proportional to  $1/V$ . Here the statement that pressure and volume are inversely proportional accomplishes more in a shorter space and with less ambiguity than the equivalent description of pressure changes *causing* contractions and expansions or of volume changes *causing* compressions or rarefactions.

A great aid to speedy understanding and manipulation of concepts is the simple fact that mathematically formulated ideas can be expressed symbolically in equations like the above. First consider the mere *convenience* of writing an equation like  $s = v_0t + \frac{1}{2}gt^2$ , instead of expressing this in a sentence. Note how extraneous word colors disappear; how easy it becomes to communicate arguments and results clearly to your fellow men; and how such equations invite drawing further conclusions about the relationship between observables. There is a good parallel here in the field of chemistry, which was immeasurably advanced in the early 19th century by the simple expedient of expressing reactions in formulas.

However, still other consequences follow from the decision of scientists to restrict and direct their attention toward the measurables and the mathematical relations between them. If the important thing about a rolling ball on an inclined plane is no longer its composition, its history in the workshop, its color, or its usefulness for the sport of bowling, if all that really matters to the particular group of investigators is the relationship  $s \propto t^2$ , then this ball ceases to be an individual entity—it equally might have been a smooth loaf of bread or a round bottle of wine—and in the long run the real ball seems totally forgotten, its place being taken by “a point mass,” an idealization abstracted from the experiment. In this sense the laws of science and the controlled experiment do not directly deal with “real bodies” but with abstractions moving in a hypothetical pure space with properties of its own, in a world which we can manipulate at will, now thinking away all air resistance, now regarding the inclined plane as perfectly smooth and straight, now changing only one aspect, perhaps the angle of inclination, and leaving all other factors untouched. As seen from the outside, this world, rendered in an esoteric language of its own and filled with seemingly grotesque simplifications and exaggerations, is quite analogous to that of the modern painter, poet, or composer.

All the features of science conspire to transform real weights in beautiful Pisa into undistinguished particles moving in idealized space. It is true that much would be lost by this transposition if that were the end of the scientific process, although even this much,

the first half, yields the unique and absorbing satisfaction of being able to reduce, order, and so to understand raw sense experience. But there is more; at no point is the contact with reality completely broken, the same rules by which the transition was made are also applicable on the return trip. This is guaranteed by the operational nature of the concepts employed in that abstract world. For example, while you were deriving the equation  $v^2 = v_0^2 + 2as$  from the other equations of motion, you may not have felt in touch with the real world of moving bodies, yet at the end the result, that equation, could again be referred to real motions. In fact—and this is most important—*our mathematical world in which the calculations could proceed is justified and taken seriously by physical science only insofar as it does yield new knowledge about the real world around us.* If it were to fail in this function, we should have to rearrange the rules within the abstract world until they finally gave a useful harvest. As it happens, the abstraction from a real ball to a point did give us laws of motion applicable not only to this ball but to all other balls, to all simple round and sliding objects, and the like.

This is indeed making a good profit in return for sacrificing temporarily the identity of the real spherical object. The philosopher Rudolph Carnap has suggested a neat analogy at this point. The symbols and equations of the physicist bear the same relation to the actual world of phenomena as the written notes of a melody do to the audible tones of the song itself. The written notes, of course, do not make sounds by themselves, and yet the sounds are not altogether lost in them. At any time one can retranslate the marks on paper into the audible melody, provided one knows the convention which relates notes to tones. Correspondingly, we might now visualize science as an arch resting on two pillars, observation and experience, with conceptualization and abstraction supported in between. The security of the arch depends on the firmness of the pillars.

A whole view of science can be developed along these lines. We find an early hint in Francis Bacon's words, "But my course and method, as I have often clearly stated and would wish to state again, is this—not to extract works from works or experiments from experiments, as an empiric, but from works and experiments to extract causes and axioms, and again from those causes and axioms new works and experiments, as a legitimate interpreter of nature." A modern view along such lines is to be seen in James B. Conant's definition of science: "Science is an interconnected series of concepts and conceptual schemes that have developed as the result of experi-

mentation and observation and are fruitful of further experimentation and observations."

**12-7 Explanation.** We must now turn to answer a voice which has been questioning a fundamental point—the statement that a physical scientist finds *explanation* by discovering mathematical laws between observables. Someone might well be saying, "I can see that this convention works in the sense that it permits one to form the laws of, say, projectile motion by means of which one can accurately direct missiles. However, I do not agree that anything is *explained* thereby. True, when Galileo pointed out that the paths of projectiles coincided with a type of geometric curve called a parabola, he did intuit an elegant and simple description of a large number of separate cases, and he may in fact thereby have discovered a simple device for predicting all kinds of trajectories. But I want to know the ultimate reason why projectiles do go in parabolic curves, and I shall remain dissatisfied with any of your explanations until you give me a picture or some intuitive necessity for the graceful arching of projectiles." Now it is certainly simple enough to explain why trajectories are parabolic, but only in terms of the laws of gravitational force acting on massive objects, these laws being themselves mathematical relationships. Our interrogator, on the other hand, does not want more of that. He might well find satisfaction with the explanation current before the late 16th century, or perhaps with some other idea, such as a supposed existence of invisible curved tubes or tracks in ether which guide projectiles.

If pushed far enough, our interrogator will have to make a startling concession: his only tools for understanding physical phenomena, his imagination's vocabulary, are pictures, allusions, and analogies involving the primitive mechanical events of everyday life. We might say he thinks predominantly with the concrete common sense of his muscles. For example, the gravitational or electrical attraction of one body for another across a vacuum is incomprehensible to him unless we allow him to imagine between those two bodies some invisible medium that somehow transmits a real mechanical pull or push from one to the other.

This feature of the human mind, this thirst for concreteness, characterizes not only the frequent preoccupation with mechanical models within science itself, but also the most primitive type of everyday explanation. We find it symbolized in the thought of the ancient people of India that the earth was supported in space on the backs



of gigantic elephants, or in the belief of the Egyptians that Osiris weighed human souls in a hand-held balance. It would indeed be surprising if the imagery of religion, poetry, and even of early science had not found its original raw material predominantly in our ordinary, direct, everyday experiences. It is also true that some of the greatest scientific contributions were made by men whose physical intuitions were just so directed. But as modern physical science has turned to problems more and more removed from the realm of common experience, it also has become necessary to enlarge the kit of tools with which to grasp and comprehend phenomena. Our discussion of the nuclear atom will later serve as a particularly striking example of the breakdown of those more naive types of understanding that insist on intuitive, visualizable explanations; the best early statement of this theme was precisely Galileo's insistence that motion was to be explained in terms of a particular mathematical law abstracted from experience.

Perhaps it is not too frivolous to hold that "to explain" means to reduce to the familiar, to establish a relationship between what is to be explained and the (correctly or incorrectly) unquestioned preconceptions. The preconceptions of the scholastics were grandiose and sweeping (natural place, final causes, celestial *vs.* terrestrial phenomena, etc.), reflecting the impressive but perhaps overly ambitious universality of their questions and explanations. The modern scientist has trained himself to hold somewhat more limited and specialized preconceptions, e.g., that nature works according to simple models or mathematical schemes. To him the law that the mathematical sum of all energies in a system can be regarded as forever constant is the satisfying explanation for a host of phenomena, whereas the same law would itself be regarded as an unexplained mystery by non-scientists. It requires training and repeated personal success in solving physical problems to be really satisfied with a mathematical answer, just as it must surely involve a strenuous program of orientation to be fully content with the rewards and the rules of behavior in any profession or, for that matter, in our own civilization.

To take one more specific example, consider the phenomenon of radioactive decay of some one isotope of radium. The experimental fact is that a given pure sample containing initially the number  $N_0$  atoms will by and by change into chemically different substances, each radium atom giving off alpha-radiation and changing first into an atom of the element radon. After a time  $t$  seconds, only  $N$  atoms

from the original quantity  $N_0$  are still unchanged (radium), as shown best in a graph (Fig. 12-1). Again, it is an experimental fact, one that was originally difficult to get at and much prized, that this graph is, for all pure radioactive substances, an exponential curve, i.e., that the equation corresponding to this behavior is of the form

$$N = N_0 e^{-0.693t/T}. \quad (12-1)$$

Here  $N$ ,  $N_0$ , and  $t$  have been defined above,  $e = 2.718$ , and  $T$  is a constant whose value depends only on the substance involved and which tells of the rapidity of the decay.

Now it is clear that as the equation given describes and summarizes the facts neatly we may refer to it as the Law of Radioactive Decay. While this empirical mathematical law is indispensable and enlightening, it does not in the least explain *why* atoms of radioactive elements explode and change into other elements. That is another and as yet largely unsolved problem. However, the solution will inevitably again be through some mathematical law, such as a formula describing the fluctuation of forces binding together the structure of the nucleus. So ultimately the observations described by Eq. (12-1) will still be explained through mathematical formulations.

Having examined the nature of concepts in physical science, we are ready to consider how these concepts are integrated in the advance of science.

### Further Reading

*Note:* A list of further reading in connection with all chapters of Part D is given at the end of Chapter 14.

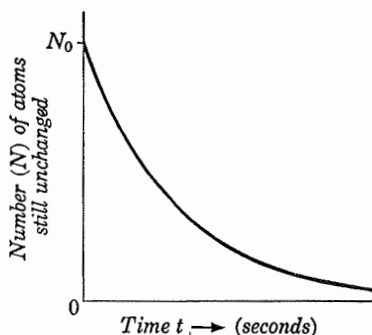


FIG. 12-1. Decay of a radioactive element.

## THE GROWTH OF SCIENCE

**13-1 The free license of creativity.** Our discussion of the nature of physical concepts may have shown that if the reason for formulating concepts is to manipulate them in mathematical laws, it is surely tempting to go one step further and to accept into science only those concepts which can be reduced to the least ambiguous of all data, numbers and measurements. The history of science would indeed furnish examples to show the great advances that followed from the formation of strictly quantitative concepts—whether the measurements were concerned with the time of descent of rolling balls, or the tabulation of population statistics, or the weights of compounds in chemical reactions, or, in the case of Pavlov's experiments, with the rate of flow of saliva in experimental animals. Should not the practicing scientist therefore deal only with ideas that correspond to strictly measurables?

With this question, we turn to the growth of science, that is, to the relationship between science and scientists. The 19th century scientist Lord Kelvin would have affirmed the last question and expressed himself in favor of such a program in the famous statement: "I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *Science*, whatever the matter may be." However, this extreme restrictive position (which at present is having a formidable influence on some thinkers in the social sciences) does justice to neither the complexity nor the fertility of the human mind, and certainly not to the needs of contemporary physical science itself—to neither scientists nor science. Quite apart from the practical impossibility of demanding of one's mind that at all times it identify such concepts as "electron" only with the measurable aspects of that construct, there are specifically two main objections: first, this position mistakes how scientists as individuals do their work and, second, it mistakes how science as a system grows out of the contributions of individuals.

Let us now examine the first of these two important points more closely. It would have been a mistake to conclude from the previous pages that practicing scientists while they are about their work do—

or even can—check the strict meaningfulness of their concepts and problems.\* As a rule an investigator is and should be unaware of such methodological questions, until some obstinate trouble develops. The operational view, where successful, is almost instinctive, and should come into the open only as a therapeutic device.

We are reminded of the case of a tightrope walker who undoubtedly is indifferent to an orderly inquiry concerning those physical laws by which he performs and survives; very likely he would fall off if he worried much about them. We all have seen, too, the deplorable results when certain contemporary artists became overly conscious of their media and methods. While a scientist struggles with a problem, there can be little conscious limitation on his free and at times audacious constructions. Depending on his field, his problem, his training and temperament, he may allow himself to be guided by a logical sequence based on more or less provisional hypotheses, or equally likely by a "feeling for things," by likely analogy, by some promising guess, or he may even follow a judicious cut-and-try procedure. The well-planned experiment is, of course, by far the most frequent one, and has the best chance of success, but men have often not even mapped out a tentative plan of attack on the problems, have instead let their enthusiasms, their hunches, and their sheer joy of discovery suggest the line of work. Sometimes, therefore, the discovery of a new effect, tool, or technique is followed by a period of trying out one or the other application in a manner which superficially seems almost playful.

The important problems of creative science do not present themselves with obvious routes to solutions, and so they differ in many fundamental ways from science problems which a student encounters in his homework or laboratory exercise. This might as well be elaborated: Above all, it is rarely known at the outset whether the problem can be solved at all with the initial conceptions or tools, or where the new complications might lurk; second, all of science has a unity and, unlike the usual student's exercise, the work of even a specialist in modern science may involve a large variety of different specific fields, sometimes in unexpected ways. An inquiry into the physical consistency of liquids, for example, may have to call upon such diverse topics as thermodynamics, physical chemistry, crystallography,

\* This is becoming more and more true if only for the simple and practical reason that in physical science nowadays there is no time for lengthy introspective speculations on one's work. This speed breeds practical, nonphilosophical scientists. They now write fewer synthetic books on large fields, but rather publish shorter specific papers.

acoustics, electronics, and even some practical metallurgy. Furthermore, the methods and results of one field in physical science constantly suggest analogous procedures in another. Consequently, the working scientist is ever alert for the slightest hints either of new difficulties or of their resolutions. He proceeds through his problem like an explorer through a jungle, sensitive to every sign with every faculty of his being. Indeed, some of the greatest theoretical scientists have stated that during those early stages of their work they do not even *think* in terms of conventional communicable symbols and words.

Only when this "private" stage is over and the individual contribution is formalized in order to be absorbed into "public" science (often called "textbook science"), only then does it begin to be really important that each step, each concept be made clear and meaningful. The difference between these two aspects of science, what we shall have occasion to call science-in-the-making or  $S_1$ , and science-as-an-institution or  $S_2$ , has thus been characterized by the American nuclear physicist H. D. Smyth, "We have a paradox in the method of science. The research man may often think and work like an artist, but he has to talk like a bookkeeper, in terms of facts, figures, and logical sequence of thought." For this reason we must not take very seriously the chronology or method outlined in the original publications of scientific results, including Galileo's own. It is part of the game to make the results in retrospect appear neatly derived from clear fundamentals until, in John Milton's phrase, "so easy it seemed/Once found, which yet unbound most would have thought/Impossible!" Years of tortuous and often wasteful effort may be hidden behind a few elegant paragraphs, just as a sculptor removes his tools and clumsy scaffolds before unveiling his work.

There is a famous example showing how dangerous it is to refuse even the provisional use of concepts until their definitions meet rigorous standards. Despite the large amount of indirect evidence for the hypothesis that matter is atomic in structure, a few prominent scientists around the turn of the century still rejected the atomic view stubbornly and vehemently as lacking "direct" confirmation. They eventually had to yield when all around them the hypothesis led to an avalanche of testable conclusions and even to a revolution of the classical concepts themselves. Although they may have been right in their position of fundamental skepticism, by waiting until the atomic idea had been operationally fortified they deprived themselves unduly long of a good mental tool—and the rest of science of their

possible additional contributions—by a doctrinal adherence to their position. It has been said very well “To set limits to speculation is treason to the future.”

In truth, most scientists today are tacitly agreed that their creative activity must be unfettered. Consider their generally uncommitted philosophic orientation. Perhaps the majority of men would defend the position that the experiences which come to them through their senses are directly occasioned by and correspond exactly to an external world, a “real” world which exists regardless of our interpretations. This view is called *Realism*, and it is in opposition to two major philosophical systems (among a host of others): first, *Idealism*, which maintains that our ideas and impressions are the only realities, and that we use those to construct convenient concepts such as chairs, houses, and meter sticks; second, *Positivism*, which does not speak of such realities at all and regards nothing as ascertainable or meaningful beyond sense data (e.g., pointer readings) or those concepts reducible to sense data. Now it might be supposed that because the established part of physical science ideally should be directly verifiable in terms of measurement, all scientists must be positivists. The true state of affairs, however, is that the great majority of scientists, while in agreement with positivism’s practical consequences for their own work, do not participate actively in the philosophical fight one way or the other. Einstein explains very well that a practicing scientist may appear to an outsider “as a type of unscrupulous opportunist: he appears as *realist*, insofar as he seeks to describe the world independent of the act of perception; as *idealist* insofar as he looks upon the concepts and theories as the free inventions of the human spirit (not logically derivable from that which is empirically given); as *positivist* insofar as he considers his concepts and theories justified only to the extent to which they furnish a logical representation of relations among sense experiences. He may even appear as *Platonist* or Pythagorean insofar as he considers the viewpoint of logical simplicity as an indispensable and effective tool of his research.” \*

**13-2 The natural selection of physical concepts.** It begins to appear that there are no simple rules to lead us to the invention of new concepts or to the discovery of new ideas, none by which to foretell if our invention or discovery, once found, is going to prove

---

\* Remarks in the collection, *Albert Einstein, Philosopher and Scientist*, page 684. Evanston, Illinois: Library of Living Philosophers (1949).

useful and durable. But we can now draw some lessons from history to see how science as a structure grows by a struggle for survival among ideas; for there is a marvelous mechanism at work in science which in time purifies the meaning even of initially confused concepts, and eventually absorbs into  $S_2$  (science-as-an-institution) anything important that may lie hidden in  $S_1$  (science-in-the-making), and so allows science to grow.

We have seen that there are two general characteristics which are shared by those concepts that have contributed most to the growth of science. First, each of the guiding concepts of physical science, acceleration, force, energy, valence, etc., has a core which is clear and unambiguous or, at any rate, which through continual application to experimental situations has attained an operational meaning that is tacitly understood and communicable. Second, by the same token, the great majority of physical concepts are quantitative, that is to say, they can be associated with numbers and measurements and therefore with manual or mathematical operations. Thus we say that a force is 15 dynes, a mass is 10 grams, a temperature 100 degrees, a valence +2, the dimensions of space 3, an atomic weight 238.07, the pH 5. This observation holds even for those concepts which superficially do not seem to take on such numerical meaning. For example, we cannot say that a body has 3, 4, or 5 units of *equilibrium*. On the other hand, equilibrium *can* be defined by saying that a body is in equilibrium when its acceleration is zero. The physical concept of "melting" can be defined as the change of state of a substance from a solid to a liquid, but this in turn has physical meaning only because the change from one state to the other can be recognized in terms of a large and abrupt change in the numerical value of measurable viscosity or hardness of the material. Consider the example of "electron." A physicist might say it is the smallest quantity of electric charge known to exist independently, but when he uses that concept he really deals with its numerical aspects, namely, that it has  $4.8 \times 10^{-10}$  units of charge,  $9.1 \times 10^{-28}$  grams mass, and so forth. The name "electron" is primarily a summary term for this whole complex of measurables.

To these two characteristics of concepts we must now add a third, for without it science would degenerate into a meaningless conglomeration of data. Evidently the chaos of experience allows us to formulate infinitely many ideas, all of them meaningful in the above sense. For example, we might arbitrarily wish to create some new concept  $X$  from observations of free fall, defining  $X$ , perhaps, by

$X = (v - v_0)/s$ . At first glance,  $X$  appears as worthy of a distinguishing name of its own and of wide acceptance, as for example the quantity  $(v - v_0)/t$ ; but pragmatically,  $X$  does not happen to appear prominently in simple laws of nature. It is not, for example, a *constant* quantity during free fall. Therefore  $X$ , though a possible, meaningful concept, actually falls by the wayside as useless and unwanted. (This  $X$  happens to be the very quantity which Galileo first and erroneously suspected of being the constant factor during free fall motion.) What makes certain concepts important, therefore, is their recurrence in a great many descriptions and laws, often in areas very far removed from the context of their initial formulation. The electron, first discovered in the current within discharge tubes rather like those now used for fluorescent fixtures, later reappeared prominently in the explanation for electric currents in wires and liquids, for photoelectricity, for thermionic phenomena (as in today's radio tubes), for radioactivity, for the emission of light from hot bodies, and much besides. *That* is the only reason and the full meaning behind the statement that scientists "believe in the reality" of electrons; the concept is needed so often and in so many ways. "The only justification for our concepts," said Einstein, "is that they serve to represent the complex of our experiences."

At the inception of an idea it is, of course, hardly possible to tell whether it will fulfill this criterion, whether it will stand the test of wide applicability and so survive in the body of science. In this sense we may say Galileo's work would have been meaningless without the successive growth of physical science. Galileo sensed, but he could not have known, that his work on mechanics would turn out to be as important as it actually did, and that his special methods and special discoveries would (with Kepler's and Newton's work) find a general significance by transcending the immediate. There is a challenge in this for each individual scientist: the science of tomorrow may well need or even depend on his own work.

The most spectacular case of such generally useful ideas is, as we saw, that of the so-called "fundamental" concepts of physics (e.g., length, time, mass). There are only a small handful of them, yet they are the building blocks from which all other concepts are constructed or derived (e.g., velocity, being the ratio of a length and a time interval). One can never cease to marvel that nature should be so reducible—here must be either the basis or the result of the scientist's guiding preconception concerning the simplicity of nature of which we have spoken. We shall meet an abundance of other



important though "derived" concepts (e.g., electric potential) which run like strong ropes through the maze of phenomena to give us support and direction under the most varied circumstances. Initially defined in some perhaps rather limited problem, they have become all-important over the years. In this evolution we find not infrequently that a concept changes its meaning as the context and the field of application widen; examples are such ideas as *force*, *energy* and *element*. This possibility of change is again suggestive of the recurring discovery that the concepts of science are not absolute, that they have meaning only in terms of what scientists can do with them.

We can now attempt a preliminary summing-up of the main points so far. We see the physical scientist as a seeker of harmonies and constancies in the jungle of experience. He aims at knowledge and prediction particularly through discovery of mathematical laws. However, this is only half the story, for science has two aspects: one (which we have called  $S_1$ ) is the speculative, creative element, the continual flow of contributions by many individuals, each working on his own task by his own usually unexamined methods, motivated in his own way, and uninterested in attending to the long-range philosophical problems of science; and the other aspect,  $S_2$ , is science as the evolving compromise, science as a growing network synthesized from these individual contributions by accepting or adopting those ideas—or even those parts of ideas—which do indeed prove meaningful and useful to generation after generation of scientists. The cold tables of physical and chemical constants, the bare equations in textbooks, are the hard core, the residue distilled from individual triumphs of insight, checked and cross-checked by the multiple testimony of general experience.

In order to illuminate this duality in terms of concrete examples, we shall now briefly study two problems, that of motivation and that of factual interpretation. This will, incidentally, also prepare the ground for the bigger tasks that follow—examining the details of the mechanism whereby  $S_2$  grows out of  $S_1$ , and explaining the vital role played by the existence of contradictory elements within science. Once an adequate distinction has been made between the several levels of meaning in the concept "science," one of the central sources of confusion in methodological discussions has been removed.

**13-3 Motivation.** When Whitehead said, "Science can find no individual enjoyment in Nature; science can find no aim in Nature; science can find no creativity in Nature," he seemed to express the

conviction of our time, and in his searching analysis even touched on the ambivalent feeling of the present age toward science—the simultaneous fear and reverence for the emotionless yet deeply exciting enterprise. Perhaps we can see now that the quotation refers to only one of the two aspects of science, the stable, not the transient one,  $S_2$  not  $S_1$ . Modern science as a structure may say nothing about purposes and aims, but the scientist himself may be enchanted largely by nonrational preoccupations (including in a certain sense even the desire to *know*, and in John Dewey's phrase, the zeal for "enjoying the doubtful").

A few direct quotations will illustrate this point, for in other years, when scientists were freer with their human secrets, the non-rational, mystical, or religious conviction was often freely acknowledged. Galileo, a pious man, looked upon the laws of nature as proof of the Deity equal to that of the Scriptures (see Section 10-2), and ever since the Pythagoreans this faith, reflected in the recent phrase "the world is divine because it is a harmony," has been a motivating theme. Almost without exception, scientists as a group have at all times, including our own, manifested the same religious convictions or lack of convictions as did other contemporary groups of educated men; if anything, their position has more often moved them to be vocal on such questions than some other, equally well-informed groups. Until about 100 years ago, the typical scientist would quite openly assert that the physical world could not be understood without fundamental theistic assumptions. In this, too, the men of the 17th century were the formulators of the fundamental concepts, and though the details of the argument have changed considerably since, its form has never been put more beautifully or stated more honestly than in Newton's description of his scientific activity as a prelude to religious knowledge (specifically, that of the "first cause," the Deity):

“. . . the main Business of natural Philosophy is to argue from Phaenomena without feigning Hypotheses, and to deduce Causes from Effects, till we come to the very first Cause, which certainly is not mechanical; and not only to unfold the Mechanism of the World, but chiefly to resolve these and such like Questions: What is there in places almost empty of Matter, and whence is it that the Sun and Planets gravitate towards one another, without dense Matter between them? Whence is it that Nature doth nothing in vain; and whence arises all that Order and Beauty which we see in the World? To what end are Comets, and whence is it that Planets move all one and the same way in Orbs concentrick, while

Comets move all manner of ways in Orbs very excentrick; and what hinders the fix'd Stars from falling upon one another? How came the Bodies of Animals to be contrived with so much Art, and for what ends were their several Parts? Was the Eye contrived without Skill in Opticks, and the Ear without Knowledge of Sounds? How do the Motions of the Body follow from the Will, and whence is the Instinct in Animals? Is not the Sensory of Animals that place to which the sensitive Substance is present, and into which the sensible Species of Things are carried through the Nerves and Brain, that there they may be perceived by their immediate presence to that Substance? And these things being rightly dispatch'd, does it not appear from Phaenomena that there is a Being incorporeal, living, intelligent, omnipresent, who in infinite Space, as it were in his Sensory, sees the things themselves intimately, and thoroughly perceives them, and comprehends them wholly by their immediate presence to himself; Of which things the Images only carried through the Organs of Sense into our little Sensoriums, are there seen and beheld by that which in us perceives and thinks. And though every true Step made in this Philosophy brings us not immediately to the Knowledge of the first Cause, yet it brings us nearer to it, and on that account is to be highly valued." (*Opticks*)

In a letter to a friend in 1692, Newton further explains his motivations for writing the *Principia*:

"When I wrote my treatise [*Principia*] about our system, I had an eye on such principles as might work with considering men for the belief of a Deity; and nothing can rejoice me more than to find it useful for that purpose."

Consider another persistent "nonrational" trend, the preoccupation with integral numbers, of which we have seen examples (Bode's Law) and shall see others. At the very outset of Galileo's historic work on the law of free fall, we find him drawing prominent attention to the fact that ". . . so far as I know, no one has yet pointed out that the distances traversed, during equal intervals of time, by a body falling from rest, stand to one another in the same ratio as the odd numbers beginning with unity." Throughout the development of physics and chemistry and to the present day we encounter this satisfaction with such simple numerical relations. Today's nuclear physicists have found that some of the characteristics of nuclei can be explained in a theory involving a few recurring numbers; these

are generally called the "magic numbers," perhaps a little less jokingly than is often realized. Examples can be multiplied indefinitely, but perhaps the following extract is sufficiently representative: The Swiss physicist, Wolfgang Pauli, on accepting the Nobel Prize of Physics, 1945, was speaking of his great teacher Sommerfeld and his attempts to form an explanation for the particular colors found in the spectra of glowing gases, when he said, "Sommerfeld however preferred . . . a direct interpretation, as independent of models as possible, of the laws of spectra in terms of integral numbers, following as Kepler once did in his investigation of the planetary system, an inner feeling for harmony . . . The series of whole numbers 2, 8, 18, 32 . . . giving the length of the periods on the natural systems of chemical elements, was zealously discussed in Munich, including the remark of the Swedish physicist Rydberg that these numbers are of the simple form  $2n^2$ , if  $n$  takes on an integral value. Sommerfeld tried especially to connect the number 8 with the number of corners of a cube."

When thus examined, the motivation of many scientists, just like the process of discovery itself, appears perhaps surprisingly "unscientific." But precisely because it is irrational, the drive toward discovery is powerful even under the most adverse conditions. How shall we explain that much great work has been done under the handicaps of extreme poverty and severe ill-health, sometimes even under the threat of imminent death? or that to this day most scientists would probably reject the ever-present lure of increased standards of living in *uncreative* positions, and instead follow their chosen work without restrictions, although with relatively few material rewards? We must recognize this symbiotic relationship: the progress of science has often depended on the almost unreasonable tenacity of its devotees and, on the other hand, the scientific activity yields a unique exhilaration and deep fulfillment. In support of this point the words of Poincaré are, as always, persuasive:

"The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful, it would not be worth knowing, and if nature were not worth knowing, life would not be worth living. Of course, I do not here speak of that beauty which strikes the senses, the beauty of qualities and of appearances; not that I undervalue such beauty, far from it, but it has nothing to do with science; I mean that profounder beauty which comes from the harmonious order of the parts and which a pure

intelligence can grasp. This it is which gives body, a structure so to speak, to the iridescent appearances which flatter our senses, and without this support the beauty of these fugitive dreams would be only imperfect, because it would be vague and always fleeting. On the contrary, intellectual beauty is sufficient unto itself, and it is for its sake, more perhaps than for the future good of humanity, that the scientist devotes himself to long and difficult labors." (*Science and Method*)

**13-4 Objectivity.** Exactly at this point we must answer a question which has been rising insistently during the last paragraphs: What of the much-vaunted objectivity of scientists? We have been told of this demand so often that we have learned to call "scientific" any inquiry that claims to be systematic and unbiased. For example, Bertrand Russell said, "The kernel of the scientific outlook is the refusal to regard our own desires, tastes, and interests as affording a key to the understanding of the world." But how can this large degree of personal involvement, just demonstrated, fail to endanger the search for objective truth? Or, since it patently does not so endanger science, how can we account for the discrepancy between the nature of motivation and of results?

The fact is that we have, by a second route, come back to the original dilemma, the apparent contradiction in function between  $S_1$  and  $S_2$ . Now we can begin the resolution, first by pointing to an extreme example. The great mathematical physicist, Sir Arthur Eddington, had such strong convictions that he once wrote a "Defense of Mysticism." Whether or not his persuasion was "meaningless" from the point of view of public science  $S_2$ , it may have been the mainspring, the deep cause, of his devoted search for truth. What is now important is that in his voluminous and distinguished scientific writings you will find no overt expression of his personal mysticism, nothing that might not equally well have been written by a gifted scientist with exactly opposite metaphysical orientation, or even one belonging to the "hard-boiled" school and quite indifferent to such questions. In modern science, personal persuasions do not intrude explicitly into the published work—not because they do not exist, but because they are expendable.

We may go a step further. In a free society, the metaphysical tenets of individual scientists, though often quite strong, are generally so varied, so vague, and so technically inept that in a sense they cancel out, made ineffectual by the lack of a basis for general accept-

ance and agreement of such tenets. It may be said that our science can be healthy only as long as this mechanism operates. Where there exists one widely accepted set of dogmas—as in some of the old scholastic universities or in a modern totalitarian state—only there can metaphysics in a scientific publication survive the scrutiny of the scientists in that field and that location, with possibly most detrimental consequences.

As contrasted with his largely unconscious motivations, the intellectual disciplines imposed on a scientist, like the form of publication of research papers itself, are now quite rigorously defined; he knows that his inclinations may well not coincide with those of his colleagues and witnesses, and that the only points he and they might agree on are sound demonstrations and repeatable experiments. The personal satisfactions must remain private and therefore incidental. We can readily see how this severe convention, in such clashing contrast with the human side of the experimenter, is a kind of sociological device ensuring the rapid evolution of science. For exactly when his enthusiasm for some result runs highest, the chances for mistakes arise. Therefore, the investigator will have to force himself to the most searching re-examination before describing his work to that most mercilessly impartial jury, his colleagues. Louis Pasteur gives us a good insight into this attitude: "When you believe you have found an important scientific fact and are feverishly curious to publish it, constrain yourself for days, weeks, years sometimes; fight yourself, try and ruin your own experiments, and only proclaim your discovery after having exhausted all contrary hypotheses. But when after so many efforts you have at last arrived at certainty, your joy is one of the greatest that can be felt by the human soul."

In attacking the qualitative and subjective science of his day, Francis Bacon held that in order to build on a sure foundation the scientist must suppress as long as possible the human element, his prejudices, premature hypotheses, his guesses and wishes. To this end Bacon outlined a procedure which prescribed the structure of scientific work so rigidly that the human element would indeed be kept at bay. In the historical context his solution was certainly not unreasonable; it might have worked, and it does seem to solve the problem of how to remain objective. But as it turned out, scientists found another, less rigorous, less safe, but far more fruitful procedure for overcoming human error, one that does so without stifling the enthusiasm and motivation of scientists, their main source of both error and truth. This discipline is *superposed* on the free creativity

of the individual; his inquiry is not controlled in its progress as if it were confined in a train moving on firmly laid out rails, but rather as though it were a rider on a horse whose obedience is assured by an occasional but firm tug on the reins.

**13-5 Fact and interpretation.** Why, you may ask, should there be any need for an imposed discipline? Is there not discipline in the very *facts* which the scientist observes? Are they not unambiguous in themselves? This raises another most interesting problem, that of the relationship between fact and interpretation.

Our very life, as well as our science, continually depends on the correct observation and classification of facts, and yet there is nothing more deceptive than facts. It is almost impossible to describe a fact without having some interpretation, some hypothesis, enter unseen—as careful cross-examination of witnesses in a courtroom often shows quite dramatically. If we follow the matter far enough, we might have to agree that the only real or basic facts are those which impress our crudest senses—the fact that we see a black or a white splotch, feel hot or cold, hear a painful or a soft sound. However, those are exactly the perceptions which Galileo called secondary quantities, and which by themselves could not conceivably have given rise to our science. Indeed, even simple animals must be able to transcend such raw observations, and thought surely does not deal with this type of isolated experience. It has been called the first law of thought that *perception* must pass into thought and knowledge via *conceptualization*. Starting from the vague impression “nowhere-then-there,” man had to invent—perhaps over a period of a thousand generations—such largely unanalyzed ideas as space, object, position, distance, time interval, motion, and velocity, before he could even think about appearances. Once more we see that scientists do have certain preconceptions and do use unexamined associations, although these may be justified by a long history of successful results. This admission seems to contradict the popular idea that the first step in science is to throw overboard all prejudgments. However, without preconceptions one cannot conceive new thoughts; therefore, what is required in science, as in other phases of life, is to be able to become aware of one’s preconceptions in an emergency, to discard those that have become meaningless, to use the others with caution, and with willingness, if found to be wrong by actual phenomena, “to pronounce that wise, ingenious, and modest sentence ‘I know it not,’” as Galileo said.

Without our almost axiomatic key concepts we should be largely deprived of intelligence and communication. Imagine a geologist describing his observation of a mountain range in terms of "pure fact." From the start he could not speak of a mountain range at all, for strictly all he sees is a change in the level of terrain. Even that implies several derived concepts, so instead of saying "There is a mountain in the distance" he would have to describe a pattern of light and darkness, and how his eyes had to focus differently upon the various features of the field of view. Recall this trivial sounding example when you encounter concepts in science (e.g., rotational inertia, electric field) which at first may seem most artificial, but which have been found indispensable in constructing conceptual schemes.

Thus our thoughts, when they would deal with observation and fact, really must manipulate concepts, constructs, ideas, with all their hidden dangers—dangers of erroneous classification, of unwarranted extrapolation or analogies, of too ambitious generalizations or too timid specializations. No wonder physical science treasures those few concepts which, through long experience, have proved to be reliable enough (e.g., length, time, mass, charge), and builds upon them the whole edifice.

The least ambiguous facts being those which involve only accounts of pointer readings and the like, physics should, as has been said, ideally start only with such descriptions. Even here is room for quarrel, for we do have to decide in advance of the experiment which are the most important significant pointer readings to be collected and described. In this sense, facts do not exist outside of the observer. Allow a layman to observe the sky through the new 200-inch telescope, and he will see nothing of interest and understand nothing he sees. But give a trained man only three good looks at a comet through a mediocre instrument, and he will call upon the theories in the field to tell you to a day how soon a comet will return, how fast it is traveling at any moment, what material it is made of, and much else. In short, the pattern we perceive when we note "a fact" is organized and interpreted by a whole system of attitudes and thoughts, memories, beliefs, and learned constructs. It is thought that gives us eyes.

Not only do brute facts alone not lead to science, a program of enthusiastic compilation of facts *per se* has more than once delayed the progress of science and, in the opinion of some observers, may be doing so even now in some contemporary research on cancer



and on cosmic rays. As Conant has pointed out, "Science advances not by the accumulation of new facts . . . but by the continuous development of new and fruitful concepts." The picture of the experimental scientist in his well-stocked laboratory subjecting matter to intensive and undirected observation is certainly absurd, though widespread. Particularly when instruments are to be used in making observations, even the first "preliminary experiments" are preceded by much nonexperimental activity, by study, thought, and calculation. Without this, the expensively obtained facts are meaningless. We conclude once more that science cannot be made by individual scientists. To understand even the simplest observations each investigator must rely on the distilled wisdom of science as an institution. When Newton modestly pictured himself as surveying the world from the shoulders of giants, he pronounced a truth applicable to every individual scientist.

**13-6 How science grows.** We may now coordinate several observations on the growth of science to gain one over-all view. The key recognition is the distinction between  $S_1$  and  $S_2$ —if you will, the difference between the man Galileo Galilei and the statement  $s \propto t^2$ . This dual view of science has an analogy in regarding the law either as *corpus juris* or as the ensemble of individual trials and juridical opinions, and in considering a tribe either as a people (a classified abstraction with recognizable stable organization, customs, etc.) or as an interacting group of individually almost unpredictable persons. Better still, compare the difference between  $S_2$  and  $S_1$  with the double interpretation one may give to animal species, either as calmly catalogued and described in zoology texts at the present state of evolution, or as exemplified by the diverse living, struggling individuals which it contains.

If we think about the last analogy more closely—and it is no more than a very helpful analogy—the mechanism of evolution for the species and for science appears to be similar in four ways: first of all, growth in both cases presupposes a mechanism of *continuity*—a species or a science can persist only if there is some stable means for handing on the structure from generation to generation in an unambiguous way. In biology the principle of continuity is found in the processes of heredity based on the highly specific nature of the genes—and in science it is identifiable with the specific operational and quantitative nature of the important concepts. Without this measure of unambiguous continuity, scientists could not communicate their work to one another and to their pupils.

Second, superposed on continuity is a mechanism of *mutation*, leaving open the constant opportunity for individual variations. In the case of biological species, of course, the process of mutation is made possible by various chemical and physical influences on the genes and on chromosome partition and recombination; in science, mutations are assured by the essential democracy of the institution and the boundless fertility of the free human mind.\*

A third mechanism is *multiplicity of effort*. To assure continuity and growth despite the low rate of occurrence of really good modifications, and in the absence of some obvious single master plan by which to proceed, science and the species alike must rely on a large number of individual attempts, from which may ultimately come those few types that are indeed useful. The uncountable bones of bygone members of the species, like the uncountable pages of the individual scientific researches of past years, are mute testimonies to the wastefulness of the process of growth, both of the zoological type and of scientific knowledge.

Finally, there is a *selection* mechanism whereby certain of the seemingly innumerable contributions and unpredictable mutations are incorporated into the continuous stream of science, a struggle among ideas not greatly different from that fight for existence in nature which allows the species to adapt itself to the changing environment. The survival of a variant under the most diverse and adverse conditions is mirrored in science by the survival of those concepts and discoveries that find usefulness in the greatest variety of further application, of those conceptual schemes that withstand the constant check against experience.

A skeptic might add that the secret of success in science is that, contrasted with art and philosophy, science has essentially an easy and clear-cut task. There are many ways to dig a well for water, for example, some no doubt very inexpert, but that problem is clear and sooner or later water will be found almost everywhere. The job itself, once formulated, is inherently a job that can be done. If science has limited itself to discovering facts and relationships, then patient minds will search for them, and in time facts will indeed be

---

\* But perhaps we must reserve a special name for those rare mutations in science, those precious occasions when, as in the work of Copernicus, Galileo, Newton, Einstein, and Bohr, a man transcends all human limitations and climbs to unsuspected heights in a "heroic effort," to use Mr. Sarton's phrase. Without this measure of heroism or prophecy or supreme dedication, the normal evolutionary progress in science might be almost as slow as that of natural evolution.

found and related. There may be much truth in this, but man is neither all-wise nor long-lived; hence what limited scientific knowledge he could gain by himself would die with him if it were not for its communicability to others, and would forever be limited by the frailties of individual minds if it were not for some selective growth. *Continuity or communicability, individual variability, multiplicity of free effort, and selective growth—these are the four related principles on which is based the success of science.*

As an immediate consequence we see now another relationship between  $S_1$  and  $S_2$ : unlike the anthill built from the tiny individual contributions of insects, science-as-an-institution does not exist *by itself* and outside the contributors, not in books or in the abstract, but in the minds of those scholars that understand and work in the field. Each physicist, each chemist, each geologist, each astronomer fulfills two functions (and it is the easy confusion between these two which has produced the common erroneous picture of him). On the one hand he is an unfettered creator in  $S_1$ , and on the other hand his mind, together with that of each of his colleagues, is the agency of transmission of  $S_2$  from one generation to the next. Consequently the content of  $S_2$ , in passing through the ages from one group of rational beings to the following, continually undergoes a twofold change: it is ever more sharpened toward communicability and unambiguity; and it is incessantly expanded by tests for continuing meaning and usefulness against the widening experience and deeper experiments.

In retrospect we can now see the signs of this mechanism of evolution in the nature of physical concepts themselves. Recall those three characteristics of scientific concepts: the first is their operational meaning, the second the preferably quantitative nature, and the third the reappearance in diverse fields of application. Each of these can be regarded as a mechanism for assuring the continuation and increase of the scientific enterprise. They aid in the unambiguous communication of problems and results, make possible unambiguous agreement (or disagreement) among different workers on facts and their interpretations, and knit together the efforts of many independent scientists, even when widely separated in subject interest, time, and space.

**13-7 Consequences of the model.** Several generally accepted features of scientific inquiry seem to be derivable from this view that science as an institution is continually filtered and modified by

its flow through the minds of original creators, and that it exhibits the stated laws of evolution. Here is a summary of a few such points.

(a) The convenience and freedom of communication is vital to the very existence of science *at each moment*. The jealous secrecy with which the alchemists hid their results doomed their efforts to stagnation and helped to delay the rise of chemistry. Again today the growing imposition of secrecy in basic research threatens to constrict the arteries of science. The right to pursue any promising lead, to publish and to exchange freely scientific information, must be vigorously defended. From this point of view it is also apparent what an important role is played by free, disinterested research in academic institutions, where a man may follow the unpredictable path of knowledge without serious constraint and interference. For in science, as in society, truth can be found only on the free market place of ideas. Only the filter of years, not the regulations of men, can extract the real worth from each contribution.

(b) The realization that science depends upon communication between its practitioners partly explains the rise of organizations among scientists. From the early 17th century there has been a rapidly accelerating multiplication of the channels of communication—societies, institutes, journals, etc. National and not infrequently international meetings are held regularly each year by the men of each specialty. Organizations have been formed for the single purpose of summarizing (abstracting), indexing, and distributing accounts of the growing avalanche of scientific developments. The scientists band together in committees and hold international congresses to decide on uniform nomenclature, definitions, and standards of measurement. They sponsor publication of monumental volumes of physical and chemical tables of constants. Above all, they are ever eager to share with one another their latest findings and difficulties.

(c) Multiplicity of effort is not just a phrase applicable to the sum of all past work, it is a staggering reality in science today. Because there exists the very real danger that an account of this kind, in presenting necessarily only some of the more spectacular and historic advances of science, may well distort the sense of proportions, let us look at least briefly at a typical set of reports on current work, selected almost at random from the standard abstracting journal that summarizes about a thousand technical papers in physics alone every month. All these were described at the same time in a recent issue of the journal:

An American mathematical physicist reports on a new way of solving a difficult type of equation, and illustrates his work by an application to typical problems involving the flow of electric currents in complex circuits.

A Swedish engineer surveys the new field of computing machines, and a scientific journal in South Africa is described as carrying an article on the possible use of such machines in solving problems in statistics.

A British journal is reported to contain a paper on astrophysics in which certain unsolved problems are reviewed and the conclusion is drawn that the desired solutions may hinge on the introduction of new physical concepts.

The French Bulletin of Astronomy has a discussion by the British Astronomer Royal on the desirability of introducing a slight change into the present set of definitions of the Second of Time in terms of specific astronomical observations.

A series of new measurements are at hand on the recently discovered high frequency radio waves, akin to "static" during sun-spot activity, received from various spots in our galaxy and beyond. One abstract reports, "The theories put forward to explain these effects are compared and it appears that a satisfactory theoretical explanation has not yet been found."

A Russian re-examines the problem of the origin of planets and their satellites, taking into account the facts of geophysics, geology, and geochemistry; and an American Nobel Prize chemist discusses the evidences for the presence of certain materials on Mars.

There are, next, hundreds of papers, including some very interesting ones from Japan and Italy, on the theory and experiments dealing with the fundamental properties of subatomic particles—their exact electric charge, behavior in electric and magnetic fields or during collisions with one another, their characteristics in cyclotrons and other big accelerators, and so forth. This is, of course, the field in the forefront of activity at the moment.

An eminent physicist has re-examined the action of electric fields on the viscosity of liquids; the effect was found to be quite small, and extremely refined experimental methods had to be devised.

A number of independent reports, including one from India and another from Germany, deal with the propagation of sound through various substances, and with the varied information on the structure of the media that can thereby be obtained. This fairly new field of research has been growing remarkably during the past few years.

and is now also attracting the interest of biology and medicine. One abstract is entitled "Detection of intracranial pathology by ultrasound."

A new measurement, claimed to be accurate to one part in 1,000,000, has been made of the speed of light by means of apparatus originally developed in wartime laboratories for use in radar. Two Americans give a new, most carefully obtained value for the wavelengths of light emitted from mercury vapor discharges; arguments are advanced in favor of adoption of a certain wavelength emission as a new standard of length.

Several advances in experimentation and theory are reported from the new and thriving field of low temperature physics. Of particular interest in this issue of the abstracting journal is a brief account of contributions to both sides of competing theories, rivaling at present for the correct interpretation of the phenomenon of superconductivity.

And there is of course a great deal more—we have at this point not even reached those abstracts dealing with crystallography, cosmic rays, molecular and atomic structure, radioactivity, and so forth. The point at issue here is, above all, the variety in topics, interests, contributors, and apparent importance of the papers. No one would venture to predict which of these contributions will be best remembered and most used thirty years hence. The great revolutions, the spectacular breaks with the past, are rare indeed. It may well be that one scientist or the other has that month planted the seeds from which a now entirely unsuspected harvest will be gathered; but the progress of science does not depend on it, as long as its pursuit is vigorous, free, and open.

(d) Our view of science as an organism obeying laws of evolution helps to explain the truism that once "a situation is ready" for a discovery or for the formulation of a conceptual scheme, it frequently is made in several places by independent workers at the same time. When we speak about some general advance and say that the time was ripe for it, we mean usually that at least three conditions obtained: there were *reasons* for looking into the given problem, that is, it now made sense to ask such a question, and it may in fact have been generally discussed; second, there were *means* for looking, i.e., the needed experimental tools were obtainable; and third, there were mental schemes for *understanding*, i.e., the needed concepts and wider conceptual systems were readily assembled. When all this happens at one time, several men may unknowingly compete; it is always possible that several scientists can expose themselves to the same

heritage, can encounter and interpret the problem in the same way and, borne by the same wave, may reach the same shore.

(e) We have had examples, and shall have others, of the very opposite phenomenon, of the rejection or long neglect that is experienced not infrequently by precisely the most startling innovators, even though their contribution later may come to be widely accepted. But if science is not simply the sum of individual contributions, if, as we saw,  $S_2$  exists not outside but within the minds of practicing scientists, then it follows that a contribution, to become part of  $S_2$ , must find its way to acceptance by and into the minds of many men. That, however, must take time, particularly if the suggestion is very radically different from the climate of opinion in the field at the moment. In Section 8-2 we saw that this is indeed in the long run desirable for the healthy evolution of science. And so it can happen (though rarely) that the contribution of a man regarded as something of a crackpot by his contemporary colleagues reappears, even if much modified, as an important idea later on when the horizons of the field have changed in the course of evolution. History clearly shows that science needed the neglected recluse in the garret as much as the rare and widely acclaimed heroes and the army of competent foot soldiers.

(f) By the same token that *freedom* in science increases the chances for both multiplicity of effort and individual variation, so also does a free science, ever productive of its own tools for further advancement, tend to become more and more a self-perpetuating or even an explosive chain reaction. As knowledge begets knowledge, it increases as though in a geometric series.

It is a source of constant wonder that once science (as we use the term) got its first real impetus in the early 17th century, it swept over Europe within a century like a long pent-up storm. The issue is, as we have seen, very complex, but one may speculate that two details contributed materially: European science had, from the 14th century on, expanded and progressed to the point where the principles of multiplicity of effort and of individual variations were finally in effective operation; in addition to this, the work of Galileo and of his contemporaries now provided attitudes and concepts which, being quantitative and operational, allowed the principle of continuity to operate more fully than before, thereby modifying the science of the time toward the self-sustaining character which it attained in the early 18th century.

(g) Since the mind of man is not divided into separate compartments, and since science exists as a complex of conceptual schemes propagated through the minds of men, we see that the one must interact with the other, that science must influence the whole culture pattern of a society and in turn be shaped by it. It is only natural that, for example, Galileo's argument on the motion of heavenly bodies, like that of his adversaries also, should be colored by and again expressed in contemporary theology, or that the impact of Newton's work on the thinking of 18th century Europe should have extended far beyond merely scientific questions. The mechanism of interaction between science and society is precisely that on which the growth of science itself depends.

(h) Let us consider one last point of universal concern, although others will suggest themselves readily. For the very same reason that one cannot foresee the course of science, one can rarely evaluate at the time of discovery whether and where it will find uses, or how important its place in pure science will be, or whether it will be twisted and abused for evil purposes by those who guide the fate of society. As one common example, consider that radioactivity was discovered rather casually by Henri Becquerel when a photographic plate became exposed by the rays from a nearby piece of uranium salt (he had put them together in a drawer in preparation for an experiment that actually was based on a plausible but erroneous notion). Consider further that Becquerel did not properly understand the meaning of his observations; that through the brave and lonely labors of the Curies the new radioactive metals were isolated and identified—an unsuspected widening of the horizon of physics and chemistry; that the new knowledge of radioactivity then stimulated a generation of workers in all sciences, and so directly and indirectly revolutionized the conceptual models of the atom, of radiation, and of chemical reactions, with very practical consequences in such widely removed studies as the treatment for certain diseases. Then recall that today radioactivity is an integral part or indispensable tool of innumerable studies quite outside the field of discovery, in biology, geology, astrophysics, metallurgy, even archeology, and many others. Surely no one will dare prescribe limits to the process of discovery!

Finally consider the great contemporary dilemma: "atomic" energy may put the philosopher's stone into every man's pocket and lead us to the happy age for which we are hardly prepared, or it may help us end the short history of our species. All our govern-



ments are simultaneously developing plans for the construction of engines for industry as well as for the destruction ("dehumanization," as one official put it recently) of hostile areas.

But as our analysis has shown us, even in this example, *to have any science at all, we must have all of it*. And though knowledge without deeper human wisdom may fail us, *less* knowledge hardly assures either wisdom or survival; moreover, one might at least doubt whether ignorance is now either possible or worth living for. For salvation and disaster are not reached by separate roads so that we may simply choose one and avoid the other; they both lie at the end of the same path, the choice being within ourselves. If we sought to solve the dilemma by fiat, by some official control of science, we should all have to submit to a censorship which, unable to foresee the result of curiosity and the future of discoveries, essentially would have to rule out all intelligent speculation. Surely, this system would guarantee the loss, along with science, of our basic freedom, while also forcing us to abandon all hope for an enlightened future. Both these results to our society would be quite comparable with the vengeance of savage conquerors.

If we look for an alternative solution, we might well try to emulate the lesson found in the growth of science itself: *the marvelous ability of men to arrive at truths by the free and vigorous exchange of intelligence*.

## CHAPTER 14

### LAWS

**14-1 Opinions on scientific procedure.** We have examined at some length the nature of concepts in physical science, and the relationship between scientists and their work. In this chapter we may now proceed to inquire whether anything positive may be said about the discovery of specific laws of science. It will be fitting and revealing to consult first the opinions of active scientists, and then to suggest a model to represent the discovery of some one hypothetical law.

We are at once reminded of Priestley's rather extreme statement regarding the lack of "sagacity" or even design in the process of discovery. And we remember also the warning not to expect uniformity of procedure. The fact is that while there has never crystallized a representative feeling among scientists in general, several types of opinions about the main elements of scientific procedure have been held prominently, and five facets have reappeared frequently enough to be of interest here. None of these by itself tells the whole story, none is meant to be in rivalry with the others. Rather we shall consider them as five supplementary aspects of the same general answer, the proportions being variables depending on the particular problem and the particular scientist.

(a) The first of these views, to continue with another quotation from Priestley's book,\* emphasizes that "More is owing to what we call *chance*, that is, philosophically speaking, to the observation of events *arising from unknown* [unsuspected] *causes*, than to any proper *design* or preconceived *theory* in the business. This does not appear in the works of those who write *synthetically* upon these subjects . . ."

Superficially one indeed cannot help but be impressed by the role of chance in scientific work. Every schoolboy is told how a falling apple directed Newton's mind toward the law of gravitation, or again how the law of simple pendulums occurred to Galileo as he was watching a swinging lamp in the Cathedral at Pisa. The initial observations leading to the epochal discoveries of the battery, many of the elements, electromagnetism, x-rays, radioactivity, and contemporary chemotherapy, were all made more or less by chance. Speaking of related sciences, W. I. B. Beveridge writes in his excellent book

---

\* Priestley, *op. cit.* (Section 12-3).

*The Art of Scientific Investigation*: "Probably the majority of discoveries in biology and medicine have been come upon unexpectedly, or at least had an element of chance in them, especially the most important and revolutionary ones."

This would dovetail with what has been said before concerning the irrational elements in the process of discovery and in the motivation for scientific work itself. However, one must look at the whole picture. The huge majority of scientific work is unspectacular in achievement as well as in genesis; if chance enters at all, it is greatly outweighed by sheer hard work. Whatever else may be true of the methods of science, they are by no means short-cuts; they operate within the matrix of a life spent in the laboratory, at the desk, or in field expeditions.

As to the role of chance in those several famous cases, there is still the unassailable dictum, "chance favors the prepared mind"; witness the obvious fact that great discoveries in science are in general not made by people previously ignorant of these fields. It is only the master of his subject who can turn to his advantage the irrational and the unsuspected. Chance observations have usually become important if they triggered off trains of thought that from all evidences might well have been arrived at more systematically by those minds. This trigger action of chance is by no means to be despised, but to interpret the stimulus correctly one must have had prior training and thought on the general subject of the discovery—a point made before when we spoke of the observation of facts.

(b) While not trying to write off Priestley's statement as erroneous, we certainly do expect other elements to enter more prominently into the process of discovery. A second such main ingredient has been described by Bridgman, who states straightforwardly, "The scientific method, as far as it is a method, is nothing more than doing one's damndest with one's mind, no holds barred . . . This means in particular that no special privileges are accorded to authority or to tradition, that personal prejudices and predilections are carefully guarded against, that one makes continued check to assure oneself that one is not making mistakes, and that any line of inquiry will be followed that appears at all promising. All of these rules are applicable to any situation in which one has to obtain the right answer, and all of them are only manifestations of intelligence."\*

Recalling what has been said before about the lack of a method,

---

\* P. W. Bridgman, *Reflections of a Physicist*, New York: Philosophical Library (1950).

about the variability among scientists and their methods, about the mechanism whereby science can grow from variable individual contributions, one can hardly fail to agree that Bridgman's statement is about as specific as we have any right to expect.

(c) In much the same vein others (among them Max Planck, the discoverer of the quantum concept) regard their work simply as an extension and refinement of common sense. Initially this view must seem incomplete, if only for the reason that scientists in dealing with the common-sense problems of day-to-day living on the whole can hardly claim to show much more aptitude than do their nonscientific neighbors. We must admit that many relations exist between science and common sense, for example, that the primitive notions and techniques spring from prescientific experience, that science had its origin to some extent in the common sense of the early craftsmen, and that even today much research proceeds on a kind of technical common-sense or cut-and-try empiricism, particularly in the applied sciences and in the neighboring continent of technology. But when we are dealing with men working in the modern physical sciences and with such conceptual schemes as nuclear theory, their common sense is "common" only insofar as their special training and endowment allow them to find the way through their problems as naturally as the craftsman doing his own work.

(d) The foregoing is supplemented by a further opinion which sees the essence of scientific procedure in some variation of Poincaré's terse characterization (duly qualified), "the scientific method consists in observing and experimenting." This widely held view has a range of interpretations and misinterpretations. In the extreme it may refer to a technique close to that advocated by Bacon—the pursuit, as we saw, of an orderly experimental search for facts while carefully keeping one's mind from the influence of prior notions and from forming early hypotheses. The available facts were to be interpreted or explained only at the very end by a process of "true" induction, that is, induction unspoiled by premature guesses and hypotheses. Now it is true that sometimes a research project in physical science reaches a stage where for months or even years theorizing on a large scale is held in check, pending the outcome of a single-minded search for specific experimental data; and it is further true that during the taking of data or the progress of a calculation scientists usually try to reduce the chances of mistakes by not becoming committed too early to a specific outcome, loving the game itself, so to speak, a great deal more than the final score. However, this

is, after all, as Bridgman pointed out, the only intelligent thing to do, and it also coincides with what we have called the discipline imposed on scientists; sooner or later someone is sure to check and repeat your experiments, observations, and calculations, and so will certainly discover faults and self-deceptions.

Whether in fact one can train one's mind to the Baconian level of control and asceticism, to completely uncommitted observation and experimentation, and whether the procedure is even fruitful as outlined, we have had reason to doubt before. Certainly the history of most scientific discoveries overwhelmingly denies these propositions. For example, Galileo's experiment with the inclined plane was surely not the result of an uncomplicated desire or plan to see what would happen to a rolling ball. On the contrary, as like as not Galileo knew rather well what he would observe. Some appreciation or foreknowledge of the type of solution to expect is almost inseparable even from the recognition of the *existence* of a problem. Galileo confessed his conviction that nature acted in the simplest possible way, hence he expected some relationship of the simplest possible kind between distances, times, and speeds. Of interest here is that some criteria are indeed needed to focus the attention on certain aspects of the total experience to the exclusion of the rest; the measurements of  $s$ ,  $t$ , etc. in a controlled experiment become valuable for an eventual solution only if these factors and no others really do allow us to determine the features of free fall. Of course, not too much is expected of one's prescience at this stage; many elements may suggest themselves which will turn out to be unessential. For example, Galileo appears to have suspected initially that the material of the falling body determines its law of fall, but he finally decided against this factor. On the other hand, the major clue may appear only at the end of a long investigation. On the whole, the problem here is similar to that of a craftsman who must have at least some conception of the final product before he decides on the medium and tools for rendering the work. The experimentalist, particularly, must design his apparatus and procedure with a fair idea of the type and magnitude of the effect to be expected, including the allowable errors and the masking secondary effects. In short, the scientist cannot just observe and experiment. He cannot just *search*; he has to search *for something*, and he must have an idea of what is to be expected—otherwise he may walk by the solution without recognizing it.

Now what has remained of the opinion "The scientific method

consists in observing and experimenting?" Much; above all, this: whether we put our theorizing at the beginning or at the end, whether we induce or deduce the important laws or use some interplay of both processes, sooner or later we must submit our work to the verdict of experience. Even that supreme example of "pure" theoretical work, the relativity theory, originated in the contemplation of some perplexing experimental results in the field of propagation of light and of charged particles; and at the end of the theoretical development Einstein returns to the observational and experimental aspects of science by discussing several possible experimental tests of the theory. Of one such proposed test he says bluntly, "If the displacement of spectral lines toward the red . . . does not exist, then the general theory of relativity will be untenable" (*Relativity*, 1920). Here we have again the persistent and striking difference between science and nonscience, this appeal to quantitative observation and experiment as last authorities.

(e) Finally, an element in scientific procedure held in highest esteem by those responsible for the spectacular achievements in contemporary theoretical physics is typified by Einstein's statement, "there is no logical way to the discovery of the elemental laws. There is only the way of intuition, which is helped by a feeling for the order lying behind the appearances." We hear from these men that they immerse themselves completely in the problem at hand, and either speculatively try to postulate or induce a principle by which the problematical situation may be derived (explained) or, in fact, allow their thought processes to be manipulated without imposing a conscious direction. Whereas others seem to go from step to step with logical certainty, these men make their large jumps accurately, as though borne by a guiding necessity. Sometimes they reveal that such solutions occur to them "in a sudden flash of insight" after long, even feverish study.\*

This method of solution by inspired postulation is rather close to the procedure advocated by Plato (e.g., in the *Phaedo*) as well as by certain scholastics, although they did not share our convention of letting the entire validity of such induced postulates depend ultimately on the verdict of quantitative experience. One can, and

---

\* If interested in this subject, turn to these thought-provoking books: *The Psychology of Invention in the Mathematical Field*, by Jacques Hadamard, Princeton University Press (1945); or *The Art of Scientific Investigation*, by W. I. B. Beveridge, New York: W. W. Norton (1950); or *Productive Thinking*, by M. Wertheimer, New York: Harpers (1945).

many do, defend the view that this was the main approach of Galileo also. In that case we may imagine that he solved the problem of free fall by postulating the principle that all bodies fall such that  $\Delta v/\Delta t$  is constant, and then deducing from this the experimentally confirmable behavior of falling and rolling bodies. In the same way it used to be a fashion, particularly among French scientists of the last century, to organize their work on a model adapted from the later Greek mathematicians—beginning with clearly stated postulates or hypotheses, then demonstrating all deducible concepts rigorously and mathematically, and finally calling in experimental confirmations as though these were secondary and more in the nature of illustrations. In that type of account, and it has by no means disappeared, the important real experiment is sometimes replaced by the “thought” experiment, an argument which says in essence, “Let us imagine this or that plausible situation. Surely you must agree that things will happen thus and thus, exactly in accord with my deductions.” Therefore Galileo would say he does not actually need to perform an experiment to see whether a stone dropped from the mast of a smoothly moving ship will strike at the foot of the mast no matter what the speed of the ship; he knows this to follow from the principle of composition of motions, previously verified for the case of simple projectile motion.

But again, however true and important the role of postulation, the unexplained appearance of a key principle in the exceptional mind does not by itself tell the whole story, particularly in the experimental physical sciences. For the moment let us keep in the foreground of our minds all these distinguished opinions on the requirements for successful inquiry: *an intuitive feeling for nature, particularly for its quantitative aspects, a habit of using one's intelligence to the utmost, a deepened common sense, a reliance first and last on observation and experiment, and the sensitivity for recognizing a favorable though unexpected turn of events.* Reviewing these opinions, we conclude that we may expect from the scientists themselves only the most general hints about their methods (what would a composer say if we asked by what method he created his concerti?), for in science, as in every field of endeavor, the most expert work is done by men who have passed beyond the need and interest to rationalize each step. In the sense that each problem has its own difficulties and each scientist his own approach, we may say that the scientific inquiry has its art as much as its methods.

Perhaps the best single terms for the five qualifications cited

above are "scientific orientation," or "scientific outlook," or "scientific attitude"; we may even venture to say that while scientific work in general is not characterized by any one single method, to a greater or less extent *scientists do share this complex of characteristic attitudes* toward their life occupation. This is shown by the ease with which men trained in one specialized field can usually adapt themselves to another field, to new problems and a new set of collaborators; and it helps to explain why we find so frequently that a great scientist has made fundamental contributions to several widely separated branches of study.

**14-2 A sequence of elements in formulations of laws.** Having allowed the several evidences to dispel the last hope for an easy formula of discovery, we must take heed not to fall into the extreme skepticism of Priestley. From our now suitably modest point of view, we may still analyze how scientific laws grow out of observations, concepts, and hypotheses. To be quite silent on this subject is to disregard or dismiss the distinguished and varied testimony of such influential writers on the philosophy of modern science as Pearson, Poincaré, Mach, Whitehead, Russell, Dewey, Morris Cohen, and more recently, to name only a few now working in this country, Bridgman, Carnap, Frank, Margenau, Nagel, Northrop, Reichenbach, Weyl, and others.

What we shall do is analyze a *hypothetical* case of the formulation of a law, without claiming that formulations *in general* do actually follow this specific pattern; for in truth the development of an established law cannot be traced accurately even in retrospect. Therefore the scheme to be discussed and its graphic representation in Fig. 14-1 are nothing more than a mnemonic device which may help us to think intelligently about the relationship between some recurring features in scientific procedure, particularly in experimental science. As an example we might keep in mind a concrete case, perhaps an interpretation of Galileo's discovery of the laws of motion of freely falling bodies and projectiles.

(a) At the top of the list of comments we must surely put the importance of the investigator's knowledge of contemporary science, including the permissible tricks of the trade and the more or less tacit assumptions such as are embodied in every intellectual enterprise. Thus, Galileo had available to him for his work on projectile motion the earlier and quite advanced inquiries of Tartaglia and Benedetti. Sometimes such prior knowledge will prove erroneous



and the assumptions misleading, but at the outset the existing conceptual schemes alone can give direction to the inquiry. They are the initial common sense from which the scientist must draw or perhaps eventually must break away.

At a later stage the investigator will undoubtedly have to go back to his books or instruments to supplement his knowledge in unfamiliar fields which turn out to touch on the developing situation. Or he may even have to supplement existing disciplines by the addition of specifically invented tools of measurements and aids of mathematics. Galileo's inventiveness, of which we have heard much, extended also to the field of mathematics. We found that other great scientists had to face the same problems—Copernicus had to invent much of his solid geometry, Newton his calculus.

(b) A second point of departure, again principally in the experimental sciences, is a first-hand acquaintance with nature through intelligent observation even before the problem has been formulated. Not only do the great "chance" observations arise in this manner, but what is probably much more important, observations provide the raw material with which to construct the great questions, and they provide the investigator with the necessary measure of non-rational insight and an almost intuitive feeling for the behavior of nature in the field of his specialty, what the significant variables are and how they interact. To return to the work of Einstein, because it is perhaps as far removed from personal participation and experimentation as is possible in physics, in his autobiographical notes we find this significant hint on the importance of first-hand experience with "raw" natural phenomena: "[At the Polytechnic Institute in Zurich] I really could have gotten a sound mathematical education. However, I worked most of the time in the physical laboratory, fascinated by the direct contact with experience."

(c) When the setting has been prepared, the "problematic situation" may present itself to the mind in an "occasion of reflection," (to use John Dewey's happy phrases), arising perhaps from some surprising chance finding, but more probably as the result either of some dissatisfaction with the degree of consistency, generality, or accuracy of the current explanation, or of some "hunch" where the rewarding new field of study lies. There is no clear biographical account to tell us why and how Galileo first decided on his study of motion, but we may gather from his remarks and from what we know of his life and work that, being dissatisfied with contemporary qualitative modes of description, he wished to find quantitative relations

in motion. Other men, notably the phenomenal 19th century experimentalist Michael Faraday, seemed to have been motivated more by imaginative guesses and intuitive feel for nature. To add one example to the evidences on the variability of motivation and scientific temperament, we are told by Niels Bohr that his historic theory of atomic structure originated in his recognition of a similarity in the *form* of two equations from different fields of physics.

(d) Intimately connected with this recognition of the existence of the problem is the foreknowledge of the type of solution to be expected, as we have seen. After the initial period of wonder, the trained but curious mind can hardly prevent its first rather vague flights of speculation and anticipation. It will suggest quite naturally some further observations and attention to specific details, and so point in a general way to a suspected answer to the problem, even though the real problem itself may on deeper thought turn out to lie quite far from the initial stimulus. For example, Wilhelm Röntgen in 1895 was led to the discovery of x-rays by noting that a sample of mineral gave off a pale light while electric current was made to pass through a discharge tube six feet away. He quickly traced this effect to the action of previously unsuspected radiation coming from the tube. But he immediately abandoned the problem of why minerals fluoresce in this new radiation (for about 20 years this problem could not even begin to be understood), and instead he turned to the far more interesting and fruitful questions, e.g., whether x-rays are a form of invisible light.

(e) This early stage of research may not only suggest new and interesting problems and some tentative hypotheses concerning the eventual solution of the problems, but it may also give rise to new concepts with which to analyze the problematic situation. To take our previous case, Röntgen found that objects are transparent to the passage of x-rays but that an increase in the thickness or the density of bodies decreased their transparency for x-rays materially. It occurred to him to compare whether the transparency of various metals is equal when the product of thickness and density is constant. It so happens that this proved not to be the case (the complexity of the problem is great) but evidently if this simple constancy had been found, very quickly a new name for the concept "thickness times density of metals" would have been assigned, equipped with a symbol of its own, and no doubt figuring prominently in the eventual solution of the problem of transparency of metals to x-rays. In the same vein, Galileo could analyze the problem of free fall in terms of

the concept "acceleration" as formulated for that very purpose.

Evidently the great concepts, like the great problems on which physics is built, do not spring ready-made from the minds of individual scientists—far from it. They develop during the inquiry or, in the case of such concepts as force, energy, etc., they actually may have behind them a long history of evolution, involving many men and many false starts. Even when we give Galileo credit for defining the crucial concept "acceleration" we should not forget that in a sense he only completed a study that went back centuries before his time.

(f) Proceeding from the initial preliminary experimentation and the first guiding ideas, there may begin to grow in the researcher's mind at first a rather modest *working hypothesis*—be it Röntgen's preliminary decision to regard x-rays as a form of ultraviolet light, or Galileo's thoughts that the discrepancies in the free fall of bodies might be regarded as the minor effect of air friction, and that a rolling ball arrives at the bottom of the incline with the same velocity that it would have if it had fallen freely through the same vertical difference of levels.

The process of inducing a hypothesis from the limited number of known facts has been compared with the unexplainable experience one may have while playing a game of chess, when one suddenly "sees through" the whole arrangement on the board and perceives in one stroke a pattern, a long chain of moves toward a successful ending. While the mechanism by which patterns are perceived and fruitful thoughts arise is at present still being debated vigorously among psychologists of various persuasions, we can at any rate point to the purpose of hypotheses in the scheme of scientific inquiry: namely, to stimulate some intelligent, directed activity, be it theoretical or experimental. To direct action, the hypothesis does not have to be perfect—we have long ceased to be surprised about the large number of correct theories first discovered while following a hypothesis that by itself proved to be wrong. But for that very reason the initial hypothesis should not be allowed to take hold of one's mind too firmly.

The very process of induction is a hazardous art, for no logical necessity exists that the predicted consequences should actually follow. To put it more bluntly, the reasoning is usually that if the product of induction, a hypothesis—or for that matter a law—is true, then facts *a, b, c, . . .* should follow. When, on testing, these facts are actually observed, the hypothesis is *perhaps* true—but there

may be numerous other and even better hypotheses which account for the same set of facts. In his book *The Scientific Outlook*, Bertrand Russell has given a rather picturesque but stimulating example to show other dangers inherent in logical processes, particularly the limitations on experimental confirmations. He asks us to imagine that the statements to be tested are "Bread is a stone" and "Stones are nourishing." An elementary syllogism informs us that on the basis of such postulates it should follow that bread is nourishing; since the conclusion can be confirmed by the simple experiment of eating bread, we might be tempted to regard the premises as correct also! The result of induction is never unique, and deduction also may easily give faulty results if only, as here, by having assumed erroneous postulates. Likewise, science wisely never pretends to explain phenomena with absolute finality. Experiments and theories may seem to succeed or fail for an equally wide variety of wrong reasons. If you seemingly confirm a hypothesis by experiment, your measurements or the inferences from the experimental data may be faulty, or a series of canceling errors may deceive you, or the experiments may be right but the hypothesis wrong except for that singular prediction. Examples for all these combinations abound in the history of science. On the other hand, when an experiment does *not* bear out a hypothesis, the effect you looked for in vain may indeed be there but too small for the instruments at your command, or you may be deceived by masking disturbances from an unrecognized source, or again there may be faults in the method or interpretation of the test.

For all this there is no preventive. Nature's full complexity and subtlety may be far greater than our imagination can grasp or our hypotheses can accommodate. There is no substitute for the uncertain processes of thought or for the precarious business of confirming hypotheses by test, but there are several tricks which have shown themselves to be very helpful in the long history of thought: to try the simplest hypotheses first, to regard no answer or result as true beyond all possible doubt, to rely on several independent checks before accepting an experimental result of importance, and constantly to re-evaluate and reformulate old knowledge in the light of new. We glimpse here again the fundamental truth that science depends on the continuous activities of generations.

Ernst Mach said, "Let us early get used to the fact that science is unfinished, variable." If we are reconciled to the fact that it is better to have a poor hypothesis than none, that our thoughts may

originally be grossly oversimplified, that they may need later and drastic modification, that sometimes several hypotheses may co-exist—then we shall not be tempted to regard as foolish all work of the past or as “certain” all work of the present. Furthermore, if we regard a hypothesis as a stimulus for action, we shall not demand of it that it be immediately and directly confirmable. In the long run, of course, hypotheses or postulates do have to lead to some testable conclusions. Even though Galileo himself had no tools to check directly the relations involved in the motion of rapidly falling bodies, his assertion that their acceleration is constant was not inherently unoperational, and has, of course, since been tested directly. (Suggest such a check. Feel free to use high-speed cameras, etc. Just how *direct* is this test?) Even the large-scale explanatory hypotheses may not lend themselves directly to experimentation, e.g., the hypothesis that the agency by which the sun pulls on all planets is the same gravitational force that pulls a stone to the earth; but at any rate we shall insist on their verification by examining the degree of agreement between observed fact and the corresponding prediction deduced from our hypothesis—in our previous example, probably some astronomical observations.

(g) As our work progresses, the initial working hypotheses will permit us to design and perform more and more specific, directed, and controlled experiments. Slowly our insight will progress, as if on two crutches—now the hypotheses helping to interpret one experiment and suggesting the next, now the experimental result modifying the hypotheses, in continual interaction. In this sense, the law of free fall (as expressed in the equation) is the end product, the one statement that best describes all the important features concerning free fall, whether executed by large bodies or small, whether applied to stone or lead, falling through large distances or small.

The appellation “law” is usually reserved in physical science for those statements of generality and trustworthiness whose descriptions involve mathematical relationships. Other words are frequently used for those statements that are mainly qualitative (e.g., the statement that the horizontal and vertical components in projectile motion add up without disturbing each other) or when they are more in the nature of a model or a hidden mechanism to aid our comprehension of phenomena (e.g., the idea that chemical reactions are explainable and predictable by considering matter as made of particles). In those cases various terms may be encountered: “principle” (e.g., superposition principle), “hypothesis” in a larger sense, or “theory”

in a limited sense (e.g., Avogadro's hypothesis, or the atomic theory), "postulate," "rule," and so on. We shall not find it important to make much of possible differences between these words, and shall usually refer to all of them as *laws* for the time being.

**14-3 The limitations of physical law.** In an important sense the word "law" is very misleading. Being built on concepts, hypotheses, and experiments, laws are no more accurate or trustworthy than the wording of the definitions and the accuracy and extent of the supporting experiments. Although the laws of nature are usually called inexorable and inescapable—partly because the word suggests the erroneous analogy with divine law—the human formulation of scientific laws certainly is neither eternally true nor unchangeable. The most certain truth about the laws of science is that sooner or later the advance of knowledge and experience will prove them to be too limited or too inaccurate. This lies in the tragic role of the law itself. It is induced from a necessarily limited number of experiments or situations, and yet, to be useful, it has to pretend—cautiously—to apply to all situations of the same kind. One example of this is the set of extensive tables which Galileo compiled for the range of projectiles. The direct experimental basis for this part of his work must have been quite small, but once he had discovered the parabolic trajectory, the essence of the law of projectile motion, he could derive and believe in consequences which had never been looked for experimentally. It follows that we must be alert for a new situation in which this ancient law will prove to break down—not only because it is an idealization never completely exemplified under ordinary conditions where air friction cannot be neglected, but also because there is, for example, no logical necessity that projectiles at extremely high speeds (as in the case of atomic particles) must obey the same laws as relatively slow cannon balls.

Here we remember the warning that all generalizations, all statements of physical law, carry a hidden "text"—including an explanation of the symbolism, a convention of how to manipulate the equations, definitions of the terms, statements of the accuracy and the range of experience supporting the law, and so forth. Almost invariably a limit is encountered sooner or later beyond which the law is no longer valid in its stated form; therefore the warning implied in the "text" should prevent continual misapplication of the law. An analogy may be found in the awareness of a party of mountain climbers concerning the limits beyond which they can put no demands

on their ropes, their gear, and their own bodies. In both cases the *precise* limits can usually not be clearly drawn until some failure brings them to attention. But whereas the climbers naturally dread such a failure, scientists are not helpless if occasionally a new phenomenon does not obey these established rules. For beyond the limits of one law may lie a hitherto undiscovered part of science—new and exciting knowledge. We shall see several great theories, such as Planck's and Bohr's, that had their beginning in some such "unexplainable" observations.

Galileo's laws of projectile motion contain such a "text." We find in his work a thorough examination of the major limitations involved; for example, he points out that the trajectory will not be parabolic for extremely large ranges because then the acceleration of gravity, being directed toward the center of our round earth, will no longer be parallel for different parts of the paths assumed in our equations. Of course this failure applies, strictly speaking, even to small trajectories, but in such cases it is by theory and test permissible to neglect it for our purposes, just as we have already neglected the effect of air resistance, rotation of the earth, variation of the value of  $g$  over the whole path, etc. Of these limitations within which our laws operate we certainly need not be ashamed; for if we were required to deal with actual motions in which these secondary factors are not negligible, we are confident that we would surely be able to discover the laws guiding those smaller effects, and add them to our calculations.

But what if a law *is* found to break down at some point? Usually one need not become too concerned, for as a rule it is possible to rejuvenate and extend the law, perhaps by redefining some of the concepts used. Yet, one must pay the price for this flexibility in lawmaking in another way. The history of physics shows exciting cases where at one time from apparently the same set of experiments two different laws or theories were devised (e.g., the inescapable classics, namely, the 19th century wave theory *vs.* the Newtonian corpuscular theory of light). Either one of these is useful as long as no further experiments can be devised for which the two theories would predict sufficiently different answers to permit a clear choice between the rivals. However, once such a decisive experiment has been found—no easy matter—and carried out, one or the other (or possibly both) of the conceptual schemes must be conceded to be incapable of dealing with the new observation.

Yet no general scientific theory or law does thereby suddenly

become completely "wrong," since it still is as useful as ever in explaining and predicting the class of all the experiments *before* the decisive one that had incriminated the old theory or law. For instance, Newton's laws of mechanics as such are not "wrong" because they alone cannot deal with the motion of subatomic particles; Newton's laws are still as valid as they ever were in the huge range of experiences with larger bodies. Here a famous analogy has been suggested: If we liken the facts to be explained to fish in a pond, then the law or theory is the net with which we make the catch. It may turn out that our particular net is not fine enough to haul in *all* fish, large and small—but it may still be quite satisfactory for supplying our daily needs. We may go even further and maintain that to be useful at all, our conceptual schemes, like our nets, *must* contain some holes; if it were otherwise, we should not be able to distinguish between the significant and the trivial, the catch and the rest of the pond.

**14-4 The content of science. Summary.** We have tried to trace out the hypothetical development of a physical law. Let us in summary survey the three main elements which make up physical science. On this, at least, there is fairly good agreement among contemporary philosophers of science, although detail and phrasing may differ from one to the next. First of all there are the *concepts* or *constructs*, like velocity, mass, chemical element, etc.—the main ideas which the particular sciences use as vocabulary. They are defined by operational definitions, the rules of correspondence between the concepts on the one hand and, on the other, what can be observed (preferably measured). For example, we defined instantaneous velocity and acceleration by specific measurements and calculations, operations which involve distances, time intervals, simple algebra. Second, there are the *relationships between the concepts*. These relations may be simple factual observations (e.g., ice under pressure melts below 0°C, a humble statement involving four concepts of which at least one, temperature, is fairly treacherous), or may be more general summaries of facts called laws, principles, and so forth (e.g., Newton's second law of motion relating the net force on a body, its mass, and its acceleration by  $F = ma$ ), or may even be larger systems relating laws to one another (theories, of which we have heard in Chapter 8). Last, although here we take it for granted, we must not forget at least to mention that part of science which contains the *grammar* for expressing, verbally or mathematically, defi-



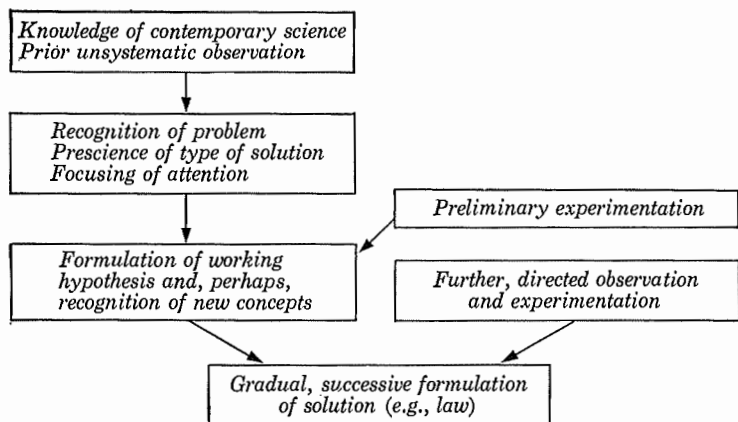


FIG. 14-1. Mnemonic diagram for relationships between recurring features in scientific procedure.

nitions of concepts and relationships between concepts, i.e., the logic of language itself (rules on how to use words like *and*, *or*), and the logic of mathematics (how to use  $+$  or  $-$ , how to add vectors, etc.).

These three parts of science are obviously so completely inter-related that any one taken by itself is quite meaningless. A physical law means nothing without the definition of the concepts and the rules of mathematics or language. A concept is useless unless it appears in relation to other concepts, or if we fail to support it with clear definitions. This has helped us to an important insight about the way science grows: an advance in any one part of science is tentative until the rest of the system absorbs the advance. For example, a newly discovered law may disturb the hitherto accepted relationships between concepts or even the usefulness of some old concept. There will be a period of rearrangement until the new discovery is fully incorporated into the evergrowing structure. The diagrammatic representation in Fig. 14-1 would be entirely misleading unless we imagined that each of the steps modified the previous conclusions, or that sometimes a few steps are tentatively skipped. The development of a physical or chemical law is perhaps analogous to the progress of a body of water down a gentle slope where there are eddies and back currents from the advancing front, and where also here and there a sudden surge reaches out to embrace and temporarily hold some territory far ahead of the main advance. Therefore we must allow that if this scheme indeed symbolizes the gradual approximation toward a law, the process may stretch over a period of gen-

erations. How fortunate for the progress of science that individual scientists do not permit themselves to become discouraged by this wide prospect of lengthy struggles—even if many do achieve their peace of mind usually by a wholehearted devotion to the narrower, day-to-day progress of their work.

### Problems

PROBLEM 14-1. What main points of difference are there between a *law of physics* (e.g., the law of projectile motion as stated in the form of our several equations) on one hand and, on the other, a *law of magic* (e.g., the belief of the Hopi Indian that the ceremonial handling of rattlesnakes increases the chances of subsequent rainfall), a *law of phrenology*, a *civil law* (law against speeding on the highway), and a *moral or divine law* (the Ten Commandments)?

PROBLEM 14-2. In *Life* magazine, issued in the spring of 1950, there appeared the following statement. Comment on the correctness or falsity of the statement and of its implications in a *short* essay.

"But the first thing to realize even about physics is its extraordinary indirectness. Physics appears to begin with very straightforward questions, but there are catches in it right from the start. For example: every high school student is told that according to Aristotle the heavier of two weights would fall faster, but that Galileo, by dropping two different weights from the leaning tower of Pisa, "proved" that Aristotle was wrong, for the weights were found to fall in exactly the same time. And yet Aristotle was right. The heavier body does fall faster, because of air resistance, which slows up the lighter body more than the heavier. Only a very little faster, it is true, and the difference may not be detectable; but since scientists claim that they use words with such great precision, it is fair to hold them to it. If you press a physicist on this point, he will readily admit that what he means is that the two bodies would fall equally fast *in a vacuum*. If you press him further, he will very reluctantly admit that nobody has ever produced a complete vacuum. And so it turns out that not the physicist but Aristotle is talking about the actual world in which we live."

PROBLEM 14-3. What exactly are the *raw observations* concerning the sun, stars, planets, satellites, etc. from which the theory of universal gravitation was derived? Try carefully to state observable facts and not to *assume* any theory in your description.

PROBLEM 14-4. Examine a law outside physics with which you are *well* acquainted (e.g., a law of economics, linguistics, biology, etc.) and decide whether or not one may trace its logical development by the same type of diagram that we have used to analyze our hypothetical physical law.

PROBLEM 14-5. Criticize these definitions: (a) The mass of an object is the measure of its quantity. (b) The range of a projectile is the hori-

zontal component of its displacement. (c) The densage of an object is its volume divided by its age. (d) The factor  $\pi$  is approximately equal to 22/7.

PROBLEM 14-6. What is meant in science by the statements: (a) Atoms exist. (b) The ether does not exist. (c) Avogadro's (or any other) hypothesis is correct. (d) The findings do not represent absolute truths.

PROBLEM 14-7. Why do *you* accept as a fact that the earth is round? that the moon is spherical? that Archimedes discovered the hydrostatic principle? Of what facts are you certain beyond all possible doubt? Why?

PROBLEM 14-8. Try to account for the fact that astronomy was one of the earliest and best-developed sciences.

PROBLEM 14-9. Assume that for his experiments Galileo had been using our most modern equipment for measuring the time and distance of freely falling bodies. Let us assume that his data had been accurate to  $10^{-6}$  sec or  $10^{-4}$  cm respectively. Now write out in a few sentences (a) what might have been the result of compiling a large set of experimental data, (b) what he actually would have had to do to discover the law of freely falling bodies by means of such data.

PROBLEM 14-10. Write a short essay on one of these topics: (a) The scientist as a human creator. (b) Science as an evolving organism. (c) Important factors in the process of scientific discovery. (d) The discovery of truths by free and vigorous exchange of intelligence. (e) The complementary aspects of experience: scientific experience and poetic experience.

PROBLEM 14-11. Write a short essay on Bridgman's extrapolation of the operational view to fields other than physical science (Section 12-4).

PROBLEM 14-12. Read again the two "deceptive" questions on page 5 and in the footnote on page 47. Then answer them in the light of Section 12-6.

PROBLEM 14-13. In the light of our analysis of the growth of science and on the basis of the prevailing social and economic structure, try to account for the fact that science flourished in Alexandria from the 4th to the 2d century B.C.; in 19th century Germany; not in 19th century America.

PROBLEM 14-14. Examine critically the analogy cited for the evolutionary processes (Section 13-6) and show where it can be extended (e.g., dangers of overspecialization) and where it breaks down.

### Further Reading

W. I. B. Beveridge, *The Art of Scientific Investigation*. New York: W. W. Norton (1950).

P. W. Bridgman, *The Logic of Modern Physics*. New York: Macmillan (1927). Chapters I, II, and the first part of III.

N. R. Campbell, *Physics, The Elements*. Cambridge: Cambridge University Press (1920).

J. B. Conant, *Science and Common Sense*. New Haven: Yale University Press (1951). Particularly Chapters 1-3.

- H. Feigl and W. Sellars, *Readings in Philosophical Analysis*. New York: Appleton-Century-Crofts (1949). Particularly pages 498-514.
- P. Frank, *Modern Science and Its Philosophy*. Cambridge: Harvard University Press (1949).
- D. Johnson, "Mysterious Craters of the Carolina Coast," *American Scientist*, January 1944. An exercise in "scientific methods."
- E. Mach, *The Science of Mechanics* (original ed. 1883). LaSalle: The Open Court Publishing Co. (1942). A venerable classic which will still reward an interested reading. First parts of Chapters 1 and 2.
- C. E. K. Mees, *The Path of Science*. New York: John Wiley (1946). Chapters III and IV.
- F. S. C. Northrop, *Logic of the Sciences and the Humanities*. New York: Macmillan (1947). Particularly Chapters I and II.
- H. Poincaré, *The Foundation of Science*, English trans. New York: Science Press (1929). Particularly Book I, Chapters 1 and 3 in "Science and Method."
- B. Russell, *Scientific Outlook*. New York: W. W. Norton (1931). Chapters II and III.
- G. Sarton, *The Study of the History of Science*. Cambridge: Harvard University Press (1936). To page 52.
- H. Shapley *et al.*, *Readings in the Physical Sciences*. New York: Appleton-Century-Crofts (1948). Pages 401-423.
- A. N. Whitehead, *Science and the Modern World*. New York: Macmillan (1925). Also Pelican *Mentor* Book M28.

## CHAPTER 15

### THE LAW OF CONSERVATION OF MATTER

**15-1 Prelude to the conservation law.** In his great work *De Rerum Natura* (*On the Nature of Things*), the Roman poet Lucretius, contemporary of Julius Caesar and Cicero, recorded and embellished the nature-philosophy of Greece's Leucippus, Democritus (both c. 450 B.C.) and Epicurus (c. 300 B.C.), and he reiterated what may be considered as one of the earliest hints of a profound general principle of science: "Things cannot be born from nothing, cannot when begotten be brought back to nothing." Everything now existing must have continual existence in past, present, and future, although form, appearance, and the like may indeed change.

Yet, there is a very considerable distance from the panegyric of Lucretius to the modern Law of Conservation of Matter—that almost axiomatic basis of much in our physical sciences—which teaches that despite changes of position, shape, phase, chemical composition, and so forth, *the total mass in a given enclosed region remains constant*. Indeed, it is quite futile to search Greek thought for direct parentage of the *specific* principles of modern physical science (except perhaps for certain astronomical ones). Lucretius' ultimate purpose, for example, was not to present a scientific text but to reduce the contemporary burden of superstition and fear through Epicurean philosophy and the rational explanation of natural phenomena. Although he documents his thesis throughout with penetrating observations from nature, he is not concerned primarily with a physical problem, but rather with a philosophical one, namely the basis for a particular type of atheism. This is made clear in one of the first pages of the *De Rerum Natura*:

"Nothing is ever gotten out of nothing by divine power. Fear in sooth takes such a hold of all mortals, because they see many operations go on in earth and heaven, the causes of which they can in no way understand, believing them therefore to be done by divine power. For these reasons when we shall have seen that nothing can be produced from nothing, we shall then more correctly ascertain that which we are pursuing, both the elements out of which every thing can be produced, and the manner in which all things are done without the hand of the gods."

**15-2 Steps toward a formulation.** Before the Law of Conservation of Matter (to be occasionally abbreviated to read LCM) could reveal itself and appear in terms meaningful to modern science, three separate developments had to take place, each in itself the work of the greatest physical scientist of his generation. First had to appear the concept of the ideally isolated system. The dominant tradition of the Greeks and scholastics was the vision of a unified, completely integrated universe. In such a scheme, the behavior of a single body is determined by its relation to the rest of the cosmos, by the necessary role it must play in the whole drama. It was not meaningful, therefore, to think of events in isolation, for example, to interpret the behavior of a single object in terms of a physical law applicable to that particular region and uninfluenced by events occurring simultaneously in the universe surrounding this region.

With Whitehead, we may regard the concept of the isolated system in part as one consequence of the work of Galileo on the motion of bodies. Specifically, Galileo's prescience of the Law of Inertia or, as we call it now, Newton's First Law of Motion, speaks of the continuing, uniform, and unaccelerated motion of any object on a straight horizontal plane in the absence of external forces; and so it invites us to map out in thought a region containing only the body in equilibrium, a region at whose boundaries all causal connections with phenomena on the outside are broken. A similar case has been discussed, in similar terms, in Chapter 4, where we wished to find the motion of one part of an Atwood's Machine or the forces on an object suspended in an accelerating lift *without* referring to the remainder of the mechanical arrangement.

Assuming, then, that there was now available the concept of an *isolated* or *closed system* by which to define the region of attention, there was next needed a criterion for measuring the quantity of matter before a conservation law could be formulated. And this was provided by Newton in the opening paragraph of the *Principia* where, speaking of the term "quantity of matter," he states, "It is the quantity that I mean hereafter under the name of body or *mass*." To know quantitatively the amount of matter in a system we need to know only its inertia (or else its weight, "for it is proportional to the weight, as I have found by experiments . . ."). With this definition in mind it was unavoidable to become aware that there is something enduring and constant about any given object—not necessarily its color or apparent size, its position or motion, or its volume or shape, or even its integrity, but rather its mass everywhere, and therefore also its weight at a given locality.

A third contribution was clearly necessary, namely, the proof that the quantity of matter in a given system or, practically speaking, the weight of some material in a closed bottle, does not change during chemical transformations. The historian of science Charles Singer\* believes that Newton's explicit proof of the constancy of weight at the same place "gave a special impetus to the rationalization of chemistry" by revealing a simple and effective touchstone for checking on quantitative changes.

To us it would seem rather simple to proceed to a direct proof of the long-suspected LCM: enclose some air and a piece of wood in a tight bottle, then focus the rays of the sun on the wood inside so that it may burn. After its transformation, the ashes and remaining gases together will be found to weigh neither more nor less than the initial components. Yet there lies a period of just about 100 years between Newton's *Principia* and the work that provided this type of final and explicit proof, the memoirs on calcination and the text *Traité Élémentaire de Chimie* (Elements of Chemistry) by Antoine Laurent Lavoisier (1743-1794), often called the father of modern chemistry (Fig. 15-1).

**15-3 Lavoisier's experimental proof.** The obstacles that had to be overcome in the interval between Newton and Lavoisier can now be stated so as to make them sound almost trivial. Lavoisier was one of the first to show conclusively that the most familiar of all chemical transformations, combustion of matter, is generally oxidation, i.e., the combination of the material with that part of the ambient air to which he gave the name *oxygen*; that therefore the gas taken from the atmosphere has to be reckoned into the over-all calculation. Before Lavoisier's time, neither the nature of gases nor that of the combustion process itself was clear enough; there was therefore little reason or facility for working with carefully measured quantities of gases, or even for carrying out the reactions in closed vessels, isolated from the rest of the chemical universe. Instead of a general law of conservation of all participating matter during chemical processes, there impressed itself upon the scientific mind first the clear, obstinate, and confusing facts that on combustion in the open some materials, like wood, would lose weight (because, as we would now say, more gases are given off than are taken on by oxidation), whereas others, like phosphorus, would conspicuously gain weight (because, in modern terms, more oxygen is fixed than

\* Charles Singer, *A Short History of Science*. Oxford: Clarendon Press (1941).



FIG. 15-1. Antoine Laurent Lavoisier (1743-1794).

vapors given up). During the reverse type of reaction, now called *reduction*, where oxygen may be given up, a similar variety of changes would occur.

Since scientific theories quite rightly and naturally tend to deal initially with concepts derived from the most conspicuous aspects of the phenomena, there had grown up in the 18th century a conceptual scheme that attempted to deal with a large number of diverse observations on combustion, including the change in the physical structure and chemical nature of the burning object, the presence of heat and of flames, the changes in the quality of the surrounding air, even the diverse changes of weight. For these purposes was invented the *phlogiston*, a substance or "principle" whose migration into or out of the transforming bodies was to account for all such observations.\*

\* For a complete account, see J. B. Conant's *Science and Common Sense*, New Haven: Yale University Press (1951), Chapter 7; also *The Overthrow of the Phlogiston Theory*, Case 2 in the Harvard Case Histories in Experimental Science, Cambridge: Harvard University Press (1950).



The phlogiston theory died within a decade or so after Lavoisier's attack on it, not only because the concept of phlogiston had tried to explain so large a range of phenomena that it failed to achieve a quantitative focus and became self-contradictory, but also because Lavoisier proved the concept to be unnecessary, and by the incontrovertible evidence of the balance could show that shifting one's attention to the total quantity of matter undergoing a chemical reaction (including the gases and vapors) would be rewarded by attaining a rigorous Law of Conservation of Matter.

A few translated excerpts from Lavoisier's text of 1789 will demonstrate something of his method in this respect, and will also hint at the relative novelty of this quantitative argument:

"The elegant experiment of Mr. Ingenhouz, upon the combustion of iron, is well known. [Then follows a description of this experiment, the burning of iron wire in a closed vessel of oxygen-enriched air.] As Mr. Ingenhouz has neither examined the change [in weight] produced on iron, nor upon the air by this operation, I have repeated the experiment under different circumstances, in an apparatus adapted to answer my particular views, as follows . . . [The modification is now given which permits accurate determination of weights.] If the experiment has succeeded well, from 100 grains [5.3 grams] of iron will be obtained 135 or 136 grains of ethiops [oxide of iron], which is an augmentation [of mass or of weight] by 35 per cent. If all the attention has been paid to this experiment which it deserves, the air will be found diminished in weight exactly to what the iron has gained. Having therefore burnt 100 grains of iron, which has required an additional weight of 35 grains, the diminution of air will be found exactly 70 cubical inches; and it will be found, in the sequel, that the weight of vital air [oxygen] is pretty nearly half a grain for each cubical inch; so that, in effect, the augmentation of weight in the one exactly coincides with the loss of it in the other."

Dozens of such experiments are described by Lavoisier, involving all types of reactions then known, and always repeating this lesson: In reactions within a closed system, the gain experienced by any one part of the system exactly counterbalances the loss by the rest of the system, or *the total quantity of matter within the system remains constant.*

Lavoisier's conceptual scheme of combustion and reduction was superior to the phlogiston theory in the form, the certainty, and the

elegance of numerical predictions, and was therefore far more favorable to scientific progress, despite the fact that now a great many observations (e.g., the presence of flames, or the change in appearance of the material) were left out of the scheme of immediate explanations. A rather similar turn of events characterized several conceptual revolutions in science, as for example the rise and fall of the caloric theory of heat, which we shall discuss at some length in Chapter 18.

Needless to say, once the problems of the composition of air and of combustion were solved, the process of oxidation assumed its place as only one special, though sometimes unusually spectacular, case of chemical reactions which all exhibit the conservation law as a general and basic proposition. As Lavoisier wrote in 1789:

“We must lay it down as an incontestable axiom, that in all the operations of art and nature, nothing is created; an equal quantity of matter exists both before and after the experiment, . . . and nothing takes place beyond changes and modifications in the combination of these elements. Upon this principle, the whole art of performing chemical experiments depends: We must [to cite one application of this principle] always suppose an exact equality between the elements of the body examined and those of the product of its analysis.”

### Problems

PROBLEM 15-1. Read Chapter 7 in Conant's *Science and Common Sense*, then discuss three chemical reactions as they would each be explained by phlogiston chemistry and by modern chemistry (e.g., the burning of coal, the calcination of a metal, the reduction of a metallic ore).

PROBLEM 15-2. Read pages 399 to 408 in *Readings in the Physical Sciences*, by H. Shapley, H. Wright, and S. Rapport (New York: Appleton-Century-Crofts, 1948), then discuss Lavoisier's experimental procedure, specifically the extent to which he used the Law of Conservation of Matter.

PROBLEM 15-3. Read the Author's Preface and Chapters III, V, and VII of Lavoisier's *Elements of Chemistry*, then discuss three reactions in which the LCM in chemical transformations is clearly demonstrated.

### Further Reading

H. Buckley, *A Short History of Physics*. London: Methuen (1927). Chapter III.

J. B. Conant, *Science and Common Sense*. New Haven: Yale University Press (1951). See Chapter VII on the development of the theory of combustion.

S. J. French, *Torch and Crucible*. Princeton: Princeton University Press (1941). A biography of Lavoisier.

W. S. Knickerbocker, *op. cit.* (Ch. 7). Chapters XIII and XVII, excerpts from Priestley and Lavoisier.

Lucretius, *On the Nature of Things*. See Book I in the edition of the Great Books Foundation, Chicago: Henry Regnery Co. (1949), also No. 750 of the Everyman's Library edition.

## THE LAWS OF CONSERVATION OF MOMENTUM

**16-1 Introduction.** After everything has been done to describe the motion of bodies accurately, to analyze the path of projectiles, the orbits of planets, and the forces needed to account for such motions, there still remains, to the most basic common sense, a deeplying and mystifying puzzle. While anyone can see intuitively that a body should move when it is somehow pushed or pulled by an agent in direct contact, it is far less clear to our primitive senses why a body should continue in motion in the absence of net forces or of any material agency in immediate contact with the body. The source of this organismic orientation is probably not far to seek; after all, we each have learned from our first days that in order to make bodies move toward us we have to take a hand in it; and to keep up the motion of anything on earth, we must be prepared to supply some effort.

To most of the Greeks and to most of the Aristotelian-Thomistic scholars of the Middle Ages, the proposition that the uniform straight-line motion of the bodies requires neither force nor agent would have been patently absurd. And even the nonuniform progress of an arrow or a bullet was explainable only by the continual commotion in the surrounding air set off at the initial moment of release. The maxim was: whatever moves must be moved by something else. Not until the days of Galileo and Newton did uniform motion without persistent and continual physical causation become generally conceivable. And significantly enough, even Newton could not free his physical universe from the need for constant association with a cause for the phenomena. Of course, to him planets no longer moved by the incessant activity of God and the angels, but at any rate motion, in His cosmic philosophy, was thought of as proceeding in the sensorium of the Deity, by virtue, so to speak, of his continual though passive attention to and knowledge of the happenings.

Actually, the attempt to liberate the idea of motion and to discuss it as an independent phenomenon goes back to the 14th century Franciscan friar William Ockham (d. 1347). The University of Paris, where he taught, became a center of thinking on mechanics through the late Middle Ages, and although it presented a minority opinion at the time, it later influenced very strongly such men as Leonardo da Vinci and Galileo himself. Ockham's contribution has been

succinctly described by Sir Edmund Whittaker in *From Euclid to Eddington* (p. 66):

“In the example of the arrow, he proposed to explain the phenomena without depending on propulsion by the air, by assuming that throughout the flight the arrow carried a certain nonmaterial cargo, so to speak (corresponding to the modern notion of momentum), the possession of which ensured the continuance of its progress. This may be regarded as the first appearance in history of any sound dynamical principle: and, taken in conjunction with his philosophical innovations, justifies us in regarding Ockham as the chief founder of modern science. His idea was carried further by his disciple Jean Buridan, who became Rector of the University of Paris in 1327, and who asserted that the cargo in question was proportional to the product of the weight of the projectile and some function of its velocity (he did not distinguish between the concepts which we call momentum and kinetic energy). His doctrines were adopted, without improvement and indeed with some deterioration, by less able successors; however, eventually they reached Galileo, Descartes, and the other physicists of the seventeenth century, who succeeded in defining momentum precisely, . . .”

Superficially it must seem that William Ockham's own suggestion is not progress but a step back toward animism. He and his pupils after him suggest that during the act of setting-in-motion, something (whether we call it momentum or, as they did, impetus, matters little here) is imparted to the body; and if this quantity then grows or diminishes within, it makes the motion itself increase or decrease. But evidently this view, while it almost endows motion with the concreteness of matter, does help to concentrate attention on the moving object itself; and when it was eventually agreed upon how to measure this “cargo,” namely, by the product of the mass  $m$  and the instantaneous velocity  $v$  of the object, then the pictorial element became unimportant. One could begin to think of motion in the abstract, and particularly of motion through a vacuum. One could forget about the question “What keeps a body in motion if not some continuous external push or pull,” and instead ask the more fruitful question “If we postulate that a moving body has at every instant a momentum  $mv$ , how does this momentum change under various interesting conditions?” Thus the original questions on motion were to be answered finally in terms of the Law of Inertia, by which an

isolated body persists in a state of constant velocity and consequently of constant momentum, or in terms of Newton's Second Law, which allows one to calculate the effect of external forces on the motion and thus on the momentum. So the scientific mind in time solved an essentially barren problem by restating it from an entirely different point of view; and historically, by and large, this has always been one of the most powerful methods toward progress in science.\*

**16-2 Definition of momentum.** The historical development leading from the early conception of momentum in the 14th century to the point which interests us now, namely, to the Law of Conservation of Momentum, lasted almost 400 years and involved a galaxy of brilliant investigations. It will not be profitable for us to examine the complex details; instead, we shall derive the essential arguments from laws known to us so far.\*\*

First of all, to repeat the definition of the concept *momentum*: it is the product of a body's mass and its instantaneous velocity, or  $mv$ ; for this we shall introduce the shorter symbol  $\mathfrak{M}$ . Being the product of a scalar quantity (mass) and a vector quantity (velocity),  $\mathfrak{M}$  is also a vector quantity, representable by an arrow pointing in the same direction as the one that signifies the corresponding velocity vector; and consequently it must also be understood that we are now dealing with *linear* momentum or *translational* momentum, in contradistinction to another type to be discussed later on.

This definition of  $\mathfrak{M}$  comes directly from the first page of Newton's *Principia*:

*Definition:* The quantity of motion [for which we use the term momentum] is the measure of the same, arising from the velocity and quantity of matter [mass] conjointly."

Newton's very first application of the concept appears later in the actual statement of his Second Law of Motion. In our Chapter 4 we rendered it in accord with more modern usage, but Newton's own words, in the Latin of the original, were

*"Lex II.* Mutationem motis proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur."

\* In this connection read H. Butterfield, *op. cit.*, Chapter I.

\*\* A partial summary of the historical development is given in E. Mach, *op. cit.*, pp. 396-421. See also pages 165-166 in Crew and De Salvio's translation of Galileo's *Dialogues Concerning Two New Sciences* for a contemporary view of projectile motion in terms of the impetus theory.

For this, the following rendition is closer to Newton's intent: The change of momentum is proportional to the motive (net) force impressed, and occurs in the direction of the straight line along which that force is impressed. In symbols, Newton would say  $F_{\text{net}} \propto \Delta(mv)$ . That this, however, for all our purposes, is equivalent to our earlier statement  $F_{\text{net}} = ma$  is seen as follows:

$$F_{\text{net}} = ma,$$

but

$$a = (v_2 - v_1)/t;$$

consequently,

$$F_{\text{net}} = m(v_2 - v_1)/t = (mv_2 - mv_1)/t = (\mathfrak{M}_2 - \mathfrak{M}_1)/t = \Delta\mathfrak{M}/t.$$

This we may rewrite variously as

$$F_{\text{net}} \cdot t = \Delta\mathfrak{M} \quad \text{or} \quad \Delta(mv), \quad (16-1)$$

or

$$F_{\text{net}} \propto \Delta\mathfrak{M},$$

which is the algebraic equivalent of Newton's statement.

*Example 1.* A 2-gm bullet is shot into a tree stump. It enters with a speed of 30,000 cm/sec and comes to rest after having penetrated 5 cm. What average force acted on the bullet during the impact?

*Solution:* From Eq. (16-1),  $F_{\text{net}} = \Delta\mathfrak{M}/t$ . Here  $\Delta\mathfrak{M}$ , or  $(mv_2 - mv_1)$ , is  $(0 - 2 \times 30,000)$  gm-cm/sec =  $-60,000$  gm-cm/sec. The negative sign here means a negative change, a decrease of momentum. The time  $t$  needed to slow down from 30,000 cm/sec to zero in 5 cm is, *assuming* uniform deceleration, found from  $s = [(v_1 + v_2)/2]t$ , and comes to  $\frac{1}{3} \times 10^{-3}$  sec by calculation. Consequently,

$$F_{\text{net}} = -60,000 / \frac{1}{3} \times 10^{-3} = -1.8 \times 10^8 \text{ dynes},$$

where the minus sign indicates that the force on the bullet (namely, the force of friction with the wood) is opposite in direction to the motion through the wood.

**PROBLEM 16-1.** (a) A 1500-kgm car is stalled on a level road. The coefficient of rolling friction between tires and road is 0.03. Use Newton's Second Law in the form of Eq. (16-1) to find how long one must push on the car with an applied force of  $1.3 \times 10^6$  dynes to make the car reach a velocity of 1 m/sec. (b) What would be the result for a 3000-lb car, an applied push of 8000 poundals, and a final velocity of 3 ft/sec? [Other data the same as in part (a).]

**PROBLEM 16-2.** A 2-lb stone is thrown at an angle of  $30^\circ$  above the level from the edge of a 400-ft cliff, at an initial speed of 60 ft/sec. What is the total change of momentum calculated from the moment of release to the instant of hitting the ground in the valley below? (*Note:* Momentum is a vector.\* Therefore  $\Delta\mathfrak{M} = \mathfrak{M}_2 - \mathfrak{M}_1$ , which is a vectorial subtraction.)

**PROBLEM 16-3.** (a) A 20-gm ball is thrown against a wall, hitting it at right angles with a speed of 10 m/sec. If the collision is perfectly elastic, the ball rebounds with unchanged speed. What is the change of momentum? (b) Solve the same problem for the same ball if it collided with the wall not perpendicularly but at an angle of  $30^\circ$  with the perpendicular direction.

**16-3 The law of conservation of (linear) momentum.** Up to this point the concept of momentum does nothing new for us. All the preceding examples and problems could have been solved directly in terms of previous knowledge. The real usefulness of the concept is to be sought in the discovery, appearing gradually by 1670 (and therefore long before the Newtonian formulation given here), that if two or more bodies experience a collision or mutual attraction or repulsion, the sum of all momenta during that interaction stays constant, or the *change* in the sum is zero, or, as Newton put it, "the quantity of motion [total momentum] suffers no change from the action of bodies among themselves." *If we consider a closed or isolated system in which  $Z$  bodies act upon one another*, the individual motions may indeed change greatly, but *the vectorial sum of all  $Z$  momenta remains constant throughout*. In symbols,

$$(\mathfrak{M}_A + \mathfrak{M}_B + \mathfrak{M}_C + \cdots + \mathfrak{M}_Z) \text{ at } t_1 \quad (16-2)$$

$$= (\mathfrak{M}_A + \mathfrak{M}_B + \mathfrak{M}_C + \cdots + \mathfrak{M}_Z) \text{ at } t_2.$$

Here, the plus sign symbolizes vectorial addition, and  $t_1$  and  $t_2$  are two different moments separated by any time interval  $\Delta t$ .

The importance, power, and generality of this *Law of Conservation of Momentum* is unimaginable without a great variety of specific examples. Consider the Eq. (16-2) as it would apply to a swarm of gas molecules. We take two "snapshots" (Fig. 16-1), perhaps a microsecond apart ( $\Delta t = 10^{-6}$  sec), and we find from measurements that the *sum* of all momentum vectors, obtained here roughly by the graphical polygon method in parts (c) and (d) of the figure, is the same in both cases, although collisions will change the individual speeds and directions of motion during  $\Delta t$ . You will notice in what follows that there are no exceptions made to the law because of the

\* All bold face letters are vector quantities.



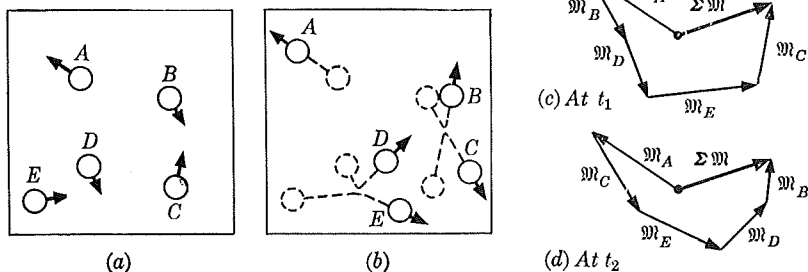


FIG. 16-1.

nature of the interactions: they can be collisions of any kind, whether elastic or inelastic; or mutual attractions or repulsions of any origin, whether magnetic, electric or gravitational, by means of springs or by explosions; or force of friction. The one and only limitation on the operation of the Law of Conservation of Momentum (to be abbreviated to LCMm) is this: that the Law in the version given applies only to a closed system, i.e., to a set of bodies that is not being acted upon by any motion-changing forces exerted in some way from outside the given region.\*

**16-4 Examples involving collisions.** The existence of the LCMm was first suspected in connection with problems involving the mechanical collision of objects.

*Example 2.* Figure 16-2, after Marcus Marci's *De Proportione Motus* (1639), may be taken to present the following (idealized) sequence of observable events in one diagram: A ball lies quietly on a table. A second ball, of equal mass, is shot horizontally toward the first, collides with it, and remains lying on the table while the other moves off with the same speed with which it was hit.

We may explain these observations (so well known to everyone acquainted with the games of croquet or billiards) directly from Newton's Laws. For during the brief time interval  $\Delta t$  of contact between the balls, the mutual forces, by the Third Law, are equally

\* Descartes thought of the LCMm, in an early form, as a necessary consequence of the "evident" fact that God had "created matter with motion and rest in its parts, and . . . thereafter conserves in the universe by His ordinary operations as much of motion and of rest as He put in it in the first creation." It is amusing but also very revealing to note that the conservation law of matter could be traced to an *atheistic* assumption, whereas here the conservation law of momentum is predicated on a *theistic* assumption.

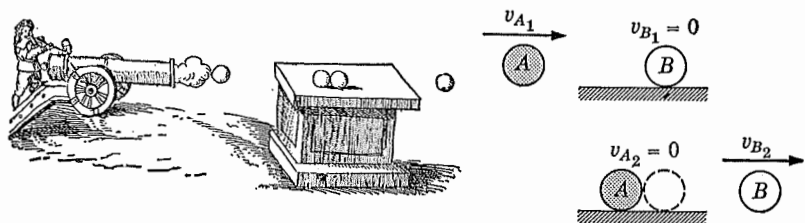


FIG. 16-2. A cannon ball collides with a stationary sphere of equal mass. (From Marcus Marci's *De Proportione Motus*.)

large though oppositely directed, therefore the decelerating (stopping) force on sphere  $A$  is equal to the accelerating force on  $B$ ; and because the two masses involved are equal, the actual deceleration of one is as large as the acceleration of the other.  $B$  gains as much speed as  $A$  loses; if  $A$  loses it all,  $B$  gains it all:  $v_{B_2} = v_{A_1}$ .

But now we shall repeat this argument in terms of the LCMm: Given are  $m_A = m_B$ ;  $v_{A_1}$ ;  $v_{B_1} = 0$ ;  $v_{A_2} = 0$ . Required:  $v_{B_2} = ?$

*Solution:* These two spheres form a closed system, for only the forces of interaction between them affect the phenomena.

The law states here  $\Sigma \mathfrak{M}$  before collision =  $\Sigma \mathfrak{M}$  after collision (where  $\Sigma$  stands for vectorial summation of the individual momentum vectors).

$$m_A v_{A_1} + m_B v_{B_1} = m_A v_{A_2} + m_B v_{B_2}. \quad (16-3)$$

(All vectors lie along the same line and in the same direction. Therefore vector addition becomes here just algebraic addition.)

On substitution from the data,

$$m_A v_{A_1} + m_B \times 0 = m_A \times 0 + m_B v_{B_2},$$

$$v_{B_2} = \frac{m_A}{m_B} v_{A_1}.$$

But

$$m_A = m_{B_1},$$

$$\therefore v_{B_2} = v_{A_1}. \quad \text{Q.E.D.}$$

**PROBLEM 16-4.** In the more general case, the masses may not be equal, but then  $v_{A_2}$  will not be zero. Nevertheless, the same type of calculation will apply. Show that then

$$v_{B_2} = \frac{m_A}{m_B} (v_{A_1} - v_{A_2}) + v_{B_1}.$$

*Example 3.* The previous example represented a perfectly elastic collision (e.g., between perfectly rigid bodies). Now note what

happens in a perfectly inelastic collision, i.e., when the colliding bodies do not separate afterwards but are wedged into each other and travel off together. For example, a car (mass  $m_A$ ) goes at speed  $v_{A_1}$  and tries to overtake a truck ( $m_B$ ) going in the same direction at speed  $v_{B_1}$ . There is a misjudgment, and a collision. The car is embedded in the rear of the truck. What is their common speed just after collision?

*Solution:* The data are  $m_A$ ;  $m_B$ ;  $v_{A_1}$ ;  $v_{B_1}$ ; and  $v_{A_2} = v_{B_2} = ?$  Here Eq. (16-3) becomes

$$m_A v_{A_1} + m_B v_{B_1} = (m_A + m_B) v_{B_2}$$

$$v_{B_2} = \frac{m_A v_{A_1} + m_B v_{B_1}}{m_A + m_B}.$$

*Example 4.* When the interacting bodies, instead of moving in the same sense, approach each other with oppositely directed momenta or, in fact, at any angle other than  $0^\circ$ , then the vector nature of momenta must be more carefully considered than we have done in the last two examples. Let two spheres, moving oppositely, hit; perhaps they are two pendulums released from different heights, meeting at the lowest point of their swing (Fig. 16-3). The collision is perfectly inelastic, they interlock;  $\mathbf{v}_{A_2} = \mathbf{v}_{B_2} = ?$  Just as we are used to finding a net force by subtracting opposing tendencies, so it is clear also that here the sum of the momenta just prior to collision is obtained by subtracting the numerical value of  $m_B v_{B_1}$  from that of  $m_A v_{A_1}$ . We still write the vector equation as before:

$$m_A \mathbf{v}_{A_1} + m_B \mathbf{v}_{B_1} = (m_A + m_B) \mathbf{v}_{A_2}$$

$$\mathbf{v}_{A_2} = \frac{m_A \mathbf{v}_{A_1} + m_B \mathbf{v}_{B_1}}{m_A + m_B}. \quad (16-4)$$

But when we now substitute values for  $\mathbf{v}_A$  and  $\mathbf{v}_B$ , we must take care, as before, to assign positive values for velocities or momenta toward, say, the right direction, negative ones to the opposite direction. If  $m_A = 3$  gm,  $m_B = 2$  gm,  $v_{A_1} = 8$  cm/sec to the right,  $v_{B_1} = 15$  cm/sec to the left, then from Eq. (16-4),

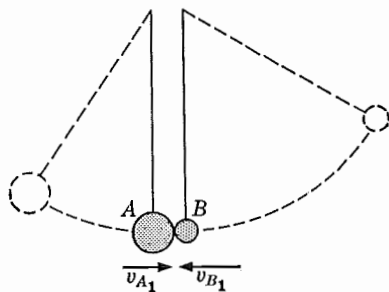


FIG. 16-3. Collision of pendulums of unequal mass.

$$v_{A_2} = \frac{(3 \text{ gm})(8 \text{ cm/sec}) + (2 \text{ gm})(-15 \text{ cm/sec})}{(3 + 2) \text{ gm}} = -1.2 \text{ cm/sec}$$

and the negative sign of the result informs us automatically that after collision both spheres move off together to the *left*. (This was, incidentally, just the type of experimentation by which Newton demonstrated the validity of the LCMm.)

**PROBLEM 16-5.** (a) Prove that two spheres of equal mass, when approaching each other with equal speeds, will also recede from each other with equal speeds after perfectly elastic collision. (b) What will happen to these spheres if they have a perfectly inelastic collision?

**16-5 Examples involving explosions.** In Chapter 4 we discussed the reaction-car experiment, useful for the determination of relative inertial masses. This, too, was a hidden example of the LCMm: It was assumed that the only unbalanced forces acting on the two cars was the mechanical push by way of the spring. This symbolizes well what happens in an explosion (Fig. 16-4). Initially all parts of the system were at rest, the momenta were zero; then mutually opposite forces separate the members of the system. We write, therefore:

$$\text{from } (\mathfrak{M}_A + \mathfrak{M}_B) \text{ before event} = (\mathfrak{M}_A + \mathfrak{M}_B) \text{ after event}$$

follows

$$0 = m_A v_{A_2} + m_B v_{B_2},$$

$$m_A v_{A_2} = -m_B v_{B_2}.$$

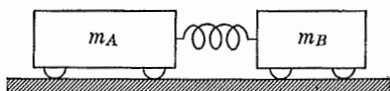


FIG. 16-4. The reaction-car experiment.

Thus we have derived the same relationship used to determine the mass  $m_B$  in terms of a standard mass  $m_A$  and observed velocities  $v_{A_2}$  and  $v_{B_2}$ .

**PROBLEM 16-6.** Prove that the same result is found for the reaction cars, no matter how strong or tight the spring is at the start.

**PROBLEM 16-7.** (a) A 1000-lb gun, initially at rest, fires a 10-lb projectile horizontally at 3000 ft/sec. If the gun is mounted on freely turning wheels, what will be its speed of recoil? (b) If the gun is, however, firmly anchored to the earth, what will be the recoil speed of gun and earth together ( $m_{\text{earth}} = 6.6 \times 10^{21}$  tons)? (c) If the barrel of the gun is 8 ft long and the projectile accelerated uniformly inside from rest to muzzle speed, what was the average force on the projectile inside the barrel? (d) What was the average force exerted at the same time on the gun by the expanding gases of combustion in the barrel?

**PROBLEM 16-8.** A sailboat is completely stalled on a lake for lack of wind. Why will it do no good to blow air into the sails by means of a

large fan fastened to the boat? (Approach this matter mathematically, using the LCMm.)

PROBLEM 16-9. Analyze in detail the method of propulsion of (a) a rocket, (b) a rotating lawn sprinkler.

*Example 5.* A gun of mass  $m_G$ , mounted on wheels, fires a projectile ( $m_P$ ) at an angle of elevation  $\theta$  (Fig. 16-5). What will be the gun's observed recoil velocity?

*Solution:* To the momentum  $\mathfrak{M}_P$  of the projectile as it leaves the gun is, of course, opposed an equally large, oppositely directed momentum in the rest of the system ( $\mathfrak{M}_S$ ). But now the case is more complex than before. The  $x$  component of  $\mathfrak{M}_S$  is, for a gun free to move horizontally, simply the horizontal recoil momentum of the gun, i.e.,  $\mathfrak{M}_x = -\mathfrak{M}_P \cos \theta$ , or, numerically,  $v_x = (m_P v_P \cos \theta) / m_G$ . However, as to the  $y$  component of  $\mathfrak{M}_S$ , the *downward* motion of the gun, on firm ground, cannot proceed without the simultaneous downward motion of the earth itself. Thus

$$\mathfrak{M}_y = -\mathfrak{M}_P \sin \theta = (m_G + m_{\text{earth}})v_y$$

or, numerically,

$$v_y = \left( \frac{m_P v_P \sin \theta}{m_G + m_{\text{earth}}} \right).$$

We must therefore supply the answer to the problem in two parts:  $v_x$ , as given above, which represents the horizontal recoil velocity, the only effect which can be observed by an onlooker standing on the earth; and  $v_y$ , as given above also, which represents the motion of gun and earth together, observable, if at all, only from a fixed point not connected with the earth (e.g., a star). Needless to say,  $v_y$  will in practice have negligibly small values.

PROBLEM 16-10. Solve problem 16-7 if that gun were mounted on wheels and fired at an angle of elevation of 55 degrees.

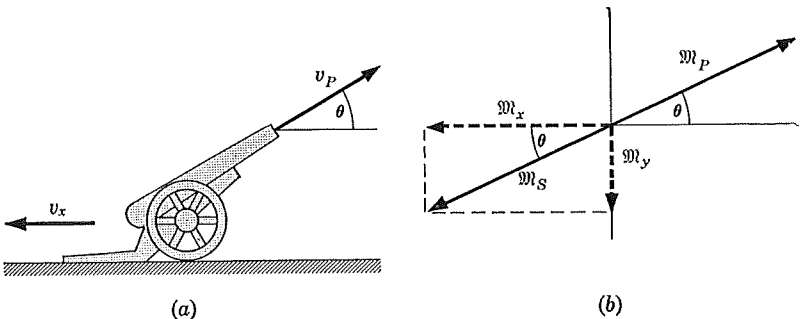


FIG. 16-5. The recoil of a gun.

**16-6 Further examples.** Following the examples in Section 16-5, we may generalize the problem of explosion by saying that the vectorial sum of all momenta in a system before the explosion is equal to the vectorial sum of the momenta of all affected members just after the explosion. It does not matter here what the nature of the "explosion," whether a relaxing spring, or expanding gases of combustion, or the release of nuclear energy, as long as different parts of the system experience mutually repelling forces. Thus, when a projectile explodes while traveling in air toward a target, its fragments, although dispersed in all directions, still continue to maintain the general forward motion,  $\Sigma \mathbf{M}$  being the same for the whole projectile before the explosion as for all the fragments just after the explosion.

One of the most interesting applications of the conservation law is to the "explosion" of a single atom during natural or induced radioactivity, and to the collision among atoms and atomic or sub-atomic particles. Lastly, we shall expand the concept of momentum to include effects involving light rays. (But note that we shall later have much more to say on the subject of nuclear physics and light.)

*Example 6.* Figure 16-6 shows the explosive break-up (fission) of one uranium atom into two about equally large fragments. The atom was at rest in the thin foil that stretches horizontally in the picture, then it was "triggered off" by a passing neutron, and the two fission products recoiled from each other in opposite directions. While they cannot, of course, be directly observed, the two fragments give themselves away indirectly by the thin streaks of fog or clouds which can be made to condense along their paths.\* It

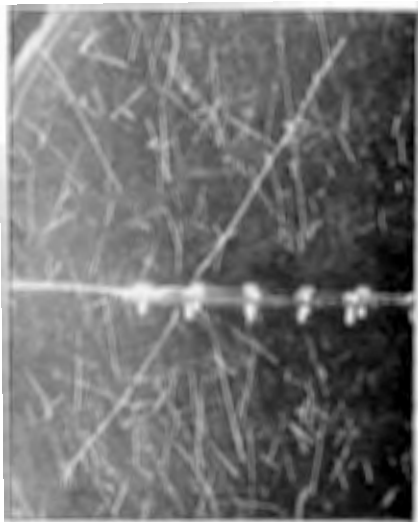


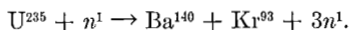
FIG. 16-6. Cloud-chamber photograph of the fission of an atom of uranium (From Bøggild, Brostrøm, and Lauritsen, *Phys. Rev.* 59, 275, 1941.)

\* The equipment for this purpose, a most important tool in contemporary research, is appropriately called a *cloud chamber*, and originated in a design devised in 1895 by the British Nobel Prize physicist C. T. R. Wilson.

is evident that the fragments are slowed down to a stop after a few centimeters by the medium through which they have to plow, but from such data as the length and density of the tracks one can deduce the speeds of separation. And from the LCMm, applied to the fission fragments, one then can get directly the ratio of their masses, an important aid in the identification of the nuclear reaction.

In this simplified account we are neglecting the contribution to momentum of minor fission products not apparent in this picture, and of the incident neutron. In the "background" one may see the result of collisions between other incident neutrons and the atmosphere in the cloud chamber.

PROBLEM 16-11. A nuclear reaction like the one just discussed might be written as follows:



By this is meant that a uranium atom (U) with a relative mass of about 235 units (1 atomic mass unit or 1 amu =  $1.6602 \times 10^{-24}$  gm) is being hit by a (slow) neutron ( $n$ , about 1 amu) which causes it to split violently into two large and three small fragments, namely, a nucleus of a Barium atom (Ba, about 140 amu) and one of a Krypton atom (Kr, about 93 amu), plus three neutrons. If we neglect all the neutrons, what will be the ratio of recoil velocities between the fission products?

Figure 16-7 presents a cloud-chamber photograph of a rather more peaceful event. An atom of oxygen in the air recoils after having been hit by an alpha-particle, i.e., a high-speed projectile having the same structure as the nucleus of a helium atom, given off

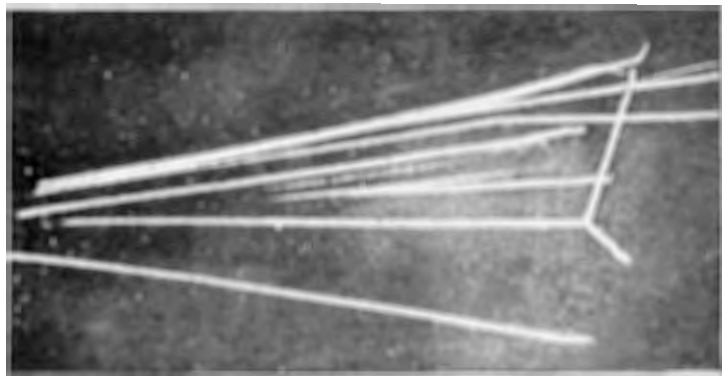


FIG. 16-7. Alpha-particles come from the left. One has collided with an atom of oxygen, which recoils to the lower right, while the alpha-particle glances off toward the top right. (After P. M. S. Blackett.)

during the natural radioactive disintegration of an element like polonium. Here we have the first example where the collision is not "head-on" but glancing.

The analysis is thereby not seriously changed, however. The  $\alpha$ -particle has about 4 amu to the oxygen atom's 16 amu (i.e., the masses are  $m_\alpha \doteq 6.6 \times 10^{-24}$  gm and  $m_O \doteq 27 \times 10^{-24}$  gm). The velocities before and after the collision are  $v_{\alpha_1}$  and  $v_{\alpha_2}$ ,  $v_{O_1}$  (= zero), and  $v_{O_2}$  (= recoil velocity). From the LCMm it follows (see Fig. 16-8) that

$$m_\alpha \mathbf{v}_{\alpha_1} = m_\alpha \mathbf{v}_{\alpha_2} + m_O \mathbf{v}_{O_2}.$$

We are already prepared to discuss the components of momenta separately, and so the vectorial equation may be rewritten for the  $x$  and  $y$  components separately:

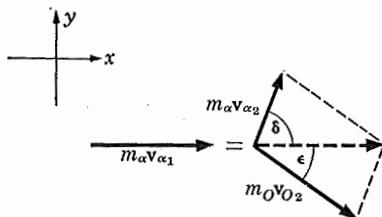


FIG. 16-8.

$$(m_\alpha v_{\alpha_1})_x = (m_\alpha v_{\alpha_2})_x + (m_O v_{O_2})_x,$$

$$(m_\alpha v_{\alpha_1})_y = (m_\alpha v_{\alpha_2})_y + (m_O v_{O_2})_y.$$

In this particular case,  $v_{\alpha_1 x} = v_{\alpha_1}$ ;  $v_{\alpha_1 y} = 0$ , by the choice of the  $x$  direction; and angles  $\delta$  and  $\epsilon$  can be obtained by measurement from the photograph. Thus

$$m_\alpha v_{\alpha_1} = m_\alpha v_{\alpha_2} \cos \delta + m_O v_{O_2} \cos \epsilon,$$

$$0 = m_\alpha v_{\alpha_2} \sin \delta + m_O v_{O_2} \sin \epsilon.$$

Evidently, from the last two equations, we could find  $m_O$  and  $v_{O_2}$  from known values of  $m_\alpha$ ,  $v_{\alpha_1}$  and  $v_{\alpha_2}$ ,  $\delta$  and  $\epsilon$ . In this way we might, in fact, identify that the recoil particle was oxygen in the first place.

**PROBLEM 16-12.** Trace (copy) Fig. 16-7 as accurately as you can for the paths discussed above. Knowing  $m_\alpha = 4$  amu,  $m_O = 16$  amu,  $v_{\alpha_1} = 5000$  miles/sec,  $v_{O_1} = 0$ , and having measured  $\delta$  and  $\epsilon$ , find the speeds of recoil of the two particles after collision.

Consider another interesting nuclear event. In Fig. 16-9, a neutron is allowed to speed into the field of view from the left after being produced elsewhere in a separate nuclear reaction. Unfortunately, neutrons do not produce tracks on which clouds settle out, and thus the direction of its entry and its initial speed are data that must be obtained in other, more indirect ways. At any rate, one such neutron collides with and enters a relatively stationary atom of



nitrogen. As a result, the latter disintegrates explosively into two unequal parts that shoot off in different directions: a heavy and slow fragment to the bottom, a faster and lighter one forward and up. The neutron itself we may imagine to have become a part of the heavier fragment. Now one can again set up the equation for the conservation of momenta, and solve, if one wishes, for the mass and speed of the incident neutron.

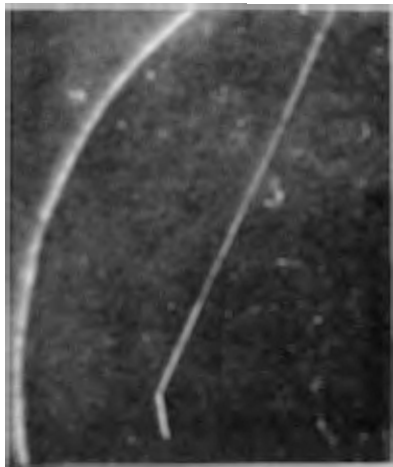


FIG. 16-9. A neutron (which leaves no track of its own) has collided with a nitrogen nucleus. The latter recoils while a proton is emitted. (After Chadwick and Feather.)

PROBLEM 16-13. Carry out this program and write the expression, in terms of the other quantities, for (a) the speed and (b) the mass of the neutron. (This exercise does not indicate the procedure or difficulty of obtaining such values in practice!)

Some of the most interesting contemporary nuclear investigations use the following simple method: The radiation whose nature is to be investigated—streams of particles from radioactive materials, or cosmic rays, or mesons produced in the big accelerators, etc.—is made to travel just inside a photographic emulsion, parallel to the surface of the plate. On developing the plate, the path of each particle may be found to have photographed itself as an exposed streak. Where such a ray encountered a suitable atomic nucleus in the emulsion, it may have triggered off a nuclear explosion on a larger scale than in the previous illustrations, as seen in Fig. 16-10. The fragments, too, photograph their own tracks, and this record can be evaluated, like that in a cloud chamber, to find from the LCMm the speeds and masses of particles. For it will evidently be true that  $\mathfrak{M}$  of the incident particle is equal to  $\Sigma \mathfrak{M}$  of all fragments.

*Example 7.* A great variety of other examples may be adduced to illustrate further the extent and power of the LCMm. Simply consider a man starting out on a walk. Before he begins, the earth and he are together at rest (as far as this problem is concerned). As soon as he moves forward, the earth below him must move back-

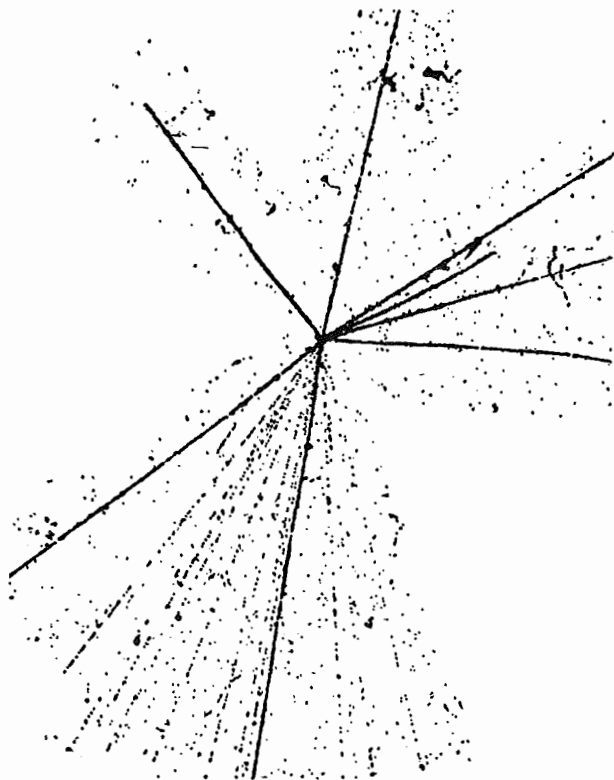


FIG. 16-10. Nuclear disintegration after cosmic-ray bombardment.  
(After C. F. Powell.)

ward, since  $\mathfrak{M}_{\text{man}} + \mathfrak{M}_{\text{earth}} = 0$ , and so  $(m\mathbf{v})_{\text{man}} = -(m\mathbf{v})_{\text{earth}}$ . To be sure, the velocity change for the earth is, as shown by calculation, extremely small, because  $m_{\text{earth}}$  is so much larger than  $m_{\text{man}}$ . On the other hand, if the surface on which the man walks is a plank on water, the relative velocities will be more nearly equal. The point, however, is the same in both cases: the force which "couples" the earth or the plank into the system of moving bodies in such a problem is *friction*.

As a force of interaction, friction differs in no way from other forces. In those previous examples that involved pendulums and reaction cars, the forces of interaction were elastic forces; in those involving recoiling guns and projectiles, they were forces set up by gases under pressure; in Example 6 they were nuclear forces. The presence of friction does not reduce the applicability of the LCMm; on the contrary, it extends it by widening any system of interacting

bodies to include all those bodies that experience mutual forces of friction. If, for instance, friction under the wheels of the reaction cars had *not* been negligible, we should have written, instead of the first equation in Section 16-5,

$$0 \text{ at start} = [(m_A \mathbf{v}_{A_2} + m_B \mathbf{v}_{B_2}) + m_{\text{earth}} \mathbf{v}_{\text{earth}}] \text{ after separation.}$$

Although this may complicate the calculation, it is still a perfectly proper application of the LCMm.

**PROBLEM 16-14.** Assume that there exists a flat interstellar platform of mass 1000 kgm, at rest as seen from some nearby star, and inhabited by a 70-kgm man. Now we see (from the star) that he walks in 3 sec a distance of 5 m, uniformly along a straight line. How far will the observed man think he walked (by his measurement along the platform)? Is this result dependent on how fast he walked? Explain. Describe the motion of the platform while he begins to walk, proceeds, and stops.

*Example 8.* Lastly, we may mention the role of another force of interaction in the operation of our Law: mutual attraction, whether gravitational, electric, or magnetic.

If we apply the conservation law to a system consisting of an apple and the earth which attracts it, their  $\Sigma \mathfrak{M}$  before the fall is zero. While the apple falls, it attracts the earth upward with a force as large as that by which it is being pulled down, and we can say

$$\mathfrak{M}_{\text{apple}} + \mathfrak{M}_{\text{earth}} = 0.$$

Apple and Earth will meet at some intermediate point which divides the total distance of initial separation in the ratio of the two masses. (Prove as a problem.) In fact, it is this very type of problem where we may find one of the earliest appreciations of the LCMm. In 1609, Kepler wrote in the *Astronomia nova*:

“If two stones were removed to any part of the world, near each other but outside the field of force of a third related body, then the two stones, like two magnetic bodies, would come together at some intermediate place, each approaching the other through a distance proportional to the mass of the other.”

Another historic example, though from 20th century physics, involves the interaction between a beam of negatively charged electrons and positively charged atomic nuclei. When a high-speed electron passes a nearby nucleus, the mutual forces of electrical attraction can deflect the former strongly from its path in exactly the

same manner as the sun affects a passing comet (Fig. 16-11). It would be expected that the change of momentum of the electron,  $\Delta\mathcal{M}_{el}$ , is compensated exactly by a corresponding and opposite  $\Delta\mathcal{M}_{nucleus}$ . On checking experimentally, we find that the latter change is too small to account for  $\Delta\mathcal{M}_{el}$ . Is this not evidence that here at last the treasured law breaks down?

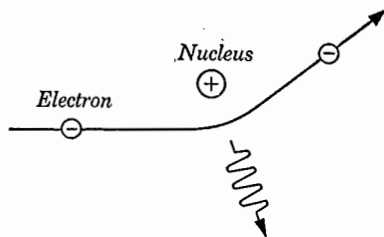


FIG. 16-11. Radiation of energy from an electron passing near a nucleus.

We should be loath to accept this position without looking first for an explanation elsewhere. We remember well that the irregularities in the motion of the planet Uranus also offered such an alternative: either to consider a failure of the old law (of universal gravitation), or else to search for an explanation of the puzzling behavior by a bold hypothesis beyond the immediately observed facts; and it was the latter course which triumphantly proved to be correct and led to the discovery of Neptune.

Here also: Does nothing else happen to carry away some momentum? During the deflection of the electron a flash of light is released in the general direction indicated, although of a frequency so far beyond the visible that instruments other than the eye are needed to detect it. If we could say that light has momentum, and assign it the missing quantity, all would be well and the law saved. But surely that is not so easily permissible! How can one conceive of momentum, *mass times velocity*, where there is nothing but disembodied wave motion, radiant energy?

While some of the detailed arguments cannot be given here and others must be delayed to Part G of this text, the main facts are these:

(i) One can no longer afford to reject from science any concept (such as the idea that light waves have momentum) just because it is not directly obvious to our mechanics-oriented thoughts. It is a dangerous limitation to insist on physical pictures. Nature has too often in the past forced us to accept what, for a time, could not possibly be imagined. After all, there was a time when it was nearly universally considered ridiculous and obviously absurd to propose that the earth might be considered to move or that the moon has mountains. Thus while one might, instead of the phrase "light has momentum," prefer to say "momentum is somehow *associated* with light waves," one must indeed say so if there is inescapable evidence that this is true.

(ii) Apart from the experiment with high-speed electrons as just described, there was other evidence that light exerts pressure and can produce changes of momentum, most spectacularly the observation that the material in the tail of comets always points away from the sun at every point along the comet's path. The suggestion that it was the sun's light which pushed the comet's tail was first made by none other than, again, Kepler. Modern and most delicate measurements around 1900 showed clearly that light waves bring with them momentum, although the quantities are ordinarily small; a good beam can do no more than keep up the rotation of a little paddlewheel mounted in a vacuum with as little friction as possible. Moreover, and most importantly, the amount of momentum in a given light beam can be predicted from simple assumptions in the theory of light, and does indeed coincide with these measured values to better than 1% despite the serious difficulties in experimentation.

(iii) The conception that a flash of light waves has associated with it a momentum became of importance subsequently in several other independent contexts; for example, in 1923 the American physicist A. H. Compton discovered that a beam of x-rays (also a

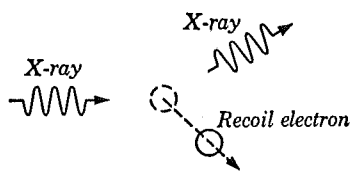


FIG. 16-12. Compton effect. The x-ray proceeds with reduced energy and longer wavelength after scattering from a recoiling electron.

form of very high-frequency light) can collide with an electron, and will bounce off with slightly changed characteristics in another direction while the electron "recoils" (Fig. 16-12) as though it were one billiard ball hit by another. Compton found that the observable momentum of the recoil electron was numerically

just equal to the change of momentum of the x-ray, as calculated by the same relation needed to fulfill the conservation law in Fig. 16-11.\* Once more the conservation law was borne out, and at the same time the nature of light itself became better understood.

**16-7 Rotational inertia.** In one sense, our concept of momentum to this point is unsatisfactory. As now defined, a body has momentum only if it travels at some velocity  $v$ , and it was understood tacitly that  $v$  refers to the motion of the center of mass (which for a regular body would mean its geometric center). But now consider a heavy

\* For this important work, which in turn had far-reaching consequences throughout modern physics, Compton was awarded a Nobel Prize (1927).

flywheel. Although it may rotate very rapidly about its axis, the velocity of translation for the wheel as a whole is zero if the machine to which it is attached is stationary. Yet, surely, to start the wheel it was necessary to apply some force  $F$ , perhaps to the rim by way of a belt, and to bring it up to rotational velocity it took time ( $t$ ). From Eq. (16-1),  $Ft = \Delta\mathfrak{M}$ , we expect that somehow the rotation, too, represents momentum.

Let us investigate briefly the fundamental facts. To begin with, consider a particle of mass  $m$  fastened to one end of a massless rod of length  $r$ . The other end of the rod is fixed at point  $O$ . A force  $F_1$  is applied to  $m$ , always perpendicular to  $r$ , as the particle accelerates in its circular path (Fig. 16-13). This acceleration  $a$  is related to  $F_1$  by  $F_1 = ma$ .

Now the product  $F_1r$  has a special significance: it is named torque,\* is assigned the letter  $\mathfrak{T}$ , and gives quantitatively the turning effect of the force  $F_1$  about a center at a perpendicular distance  $r$ . Evidently,

$$F_1r \equiv \mathfrak{T} = mar. \quad (16-5)$$

By Eq. (5-10),

$$a = \alpha r.$$

Thus

$$\mathfrak{T} = (mr^2)\alpha. \quad (16-6)$$

This last equation is for rotational motion what  $F = ma$  is for translational motion. Let us compare the three terms separately:

(a) Here it is the torque  $F_1r$ , not  $F_1$  alone, which determines the motion of the particle. A given force  $F_1$  applied at any other place along a line to center  $O$  will not produce the same rotational effect on  $m_1$ ; for example,  $F_1$  acting at  $O$  itself ( $r = 0$ ) produces no rotation at all.

(b) Whereas in translational problems the linear acceleration  $a$  is the quantity of interest, we have argued before that in rotational problems  $\alpha$  is the corresponding fundamental quantity.

(c) Whereas the inertia in linear motion is simply  $m$ , the analogous quantity in the motion now being investigated is by our derivation the term  $(mr^2)$ . A simple illustration will make this more plausible.

\* Or often the "moment" of the force, although we shall not use that term.

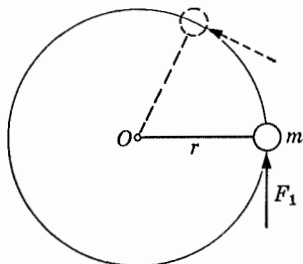


FIG. 16-13.

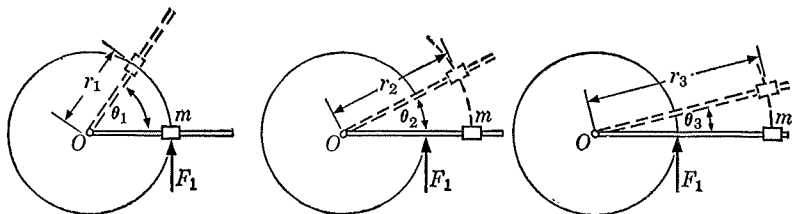


FIG. 16-14. For a given mass and a given force, the angular acceleration depends on the geometry of the system.

Figure 16-14 represents an arrangement of the same type as in Fig. 16-13, but now the mass  $m$  can be placed at three different positions along the rod. For a constant *torque*, i.e., for the same force  $F_1$  applied at the same perpendicular distance from center  $O$ , the angle  $\theta$  covered in the same initial time interval, and therefore the angular acceleration  $\alpha$  of the same mass  $m$ , is very different in these cases. The sluggishness of matter in rotational motion thus is seen to depend not only on mass  $m$  but also on the geometry of the situation. The quantity  $mr^2$ , being the ratio of torque to angular acceleration, is useful for describing this resistance of a point mass to any increase in its angular velocity at a given torque, and therefore the term is given the name *rotational inertia* (letter symbol  $I$ ):

$$I_{\text{particle}} = mr^2. \quad (16-7)$$

The units for  $I$  are evidently  $\text{gm-cm}^2$  or  $\text{lb-ft}^2$ , etc.; and Eq. (16-6) can now be rewritten in the more usual form

$$\mathfrak{T} = I_{\text{particle}} \times \alpha. \quad (16-8)$$

**PROBLEM 16-15.** In comparison with the distance from the sun, the earth itself is a point mass. What is its rotational inertia in its yearly orbit about the sun?

But the motion of point masses is rather academic. If we are dealing with an actual body, a hoop or thin ring of mass  $M$  which rotates about its own center  $O$ , each small part of the hoop, revolving at distance  $r$  and having a fractional mass  $m$ , has its own rotational inertia  $mr^2$  (see Fig. 16-15). If all such individual rotational inertias are added together, we obtain the value of  $I$  for the whole object.

$$I_{\text{hoop}} = m_1 r^2 + m_2 r^2 + \dots = \Sigma m r^2 = M r^2,$$

where  $M = m_1 + m_2 + \dots =$  total mass of the hoop.

If we wish to investigate the motion of a flat disk or solid wheel,

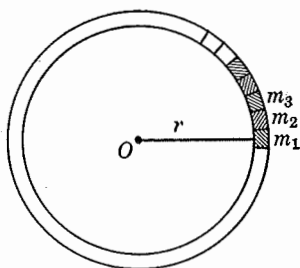


FIG. 16-15.

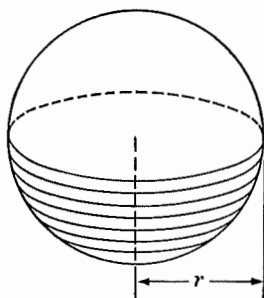


FIG. 16-17.

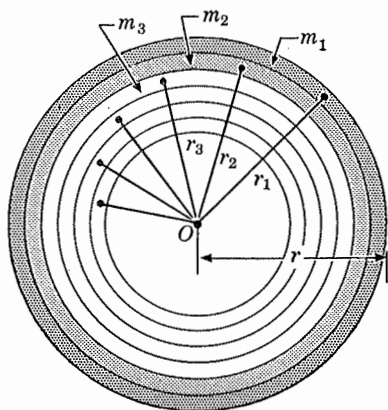


FIG. 16-16.

spinning about its own axis, we may consider the disk to be a series of concentric rings (Fig. 16-16), so that

$$I_{\text{disk}} = m_1 r_1^2 + m_2 r_2^2 + \dots$$

This sum we may evaluate by appropriate mathematical procedures to be  $I_{\text{disk}} = \frac{1}{2} M r^2$ , where  $M$  = mass of whole disk,  $r$  = its actual radius. A spinning solid sphere, in turn, is geometrically equivalent to a pile of thin disks of various radii, and its rotational

inertia is  $I_{\text{sphere}} = \frac{2}{5} M r^2$  (see Fig. 16-17). Bodies of shapes other than these will have correspondingly different values for  $I$  in terms of their masses and dimensions. But in every case the torque  $\mathfrak{T}$  needed to produce an angular acceleration  $\alpha$  is given by the same fundamental equation:

$$\mathfrak{T} = I\alpha. \quad (16-9)$$

**PROBLEM 16-16.** Three different objects have the same mass (1 kgm) and the same diameter (30 cm). One is a hoop, the other a flat disk, the third a solid sphere. What are the three values for rotational inertia? Make plausible by physical argument the difference in  $I$ , although the mass and maximum dimensions are the same for all three.

**PROBLEM 16-17.** (a) Consider the earth as a sphere spinning about its own axis only, and find its rotational inertia. (b) Comparison with problem 16-15 reveals that one may also assign another rotational inertia to the earth, *viz.*, that for its orbital motion. How can a body have two different rotational inertias?



**16-8 Angular momentum.** Our reason for investigating the concept of rotational inertia is that it is necessary for formulating another concept of great usefulness, namely, angular momentum; and this in turn, as we shall see at once, is a prerequisite for formulating another fundamental law of nature. So each step is really fully justified only in retrospect, in the context of the next higher level, and this is what makes it sometimes hard to appreciate the full beauty of a concept when it is first introduced. The same problem exists, of course, in every endeavor; if one wishes to play an instrument enjoyably, one must have practiced scales; and to reap a rich harvest, one must have tilled the soil.

From the general equation for accelerated rotational motion about a center,  $\mathfrak{T} = I\alpha$ , it follows for such cases, where the angular acceleration is constant,

$$\mathfrak{T} = I \left( \frac{\omega_2 - \omega_1}{t} \right),$$

$$\text{or} \quad \mathfrak{T} \times t = I\omega_2 - I\omega_1; \quad (16-10)$$

$$\text{in generalized form,} \quad \mathfrak{T} \times t = \Delta(I\omega). \quad (16-11)$$

We recognize this to be completely analogous to

$$F_{\text{net}} \times t = mv_2 - mv_1,$$

which was the meaning of Eq. (16-1). There, the left-hand side was the product of the net force and the time of its application (a product which technically has the name *impulse*), and the right-hand side was the corresponding change of (linear) momentum. Now that we are considering the dynamics of rotational motion, the left-hand side of Eq. (16-10) is the product of the torque times the time of its application, the right-hand side a term that might be called a change in *angular momentum*. By analogy with the terminology for translational motion along a straight line, the angular momentum of rotating or revolving objects is defined as  $I\omega$ , where  $I$  is the rotational inertia of the object with respect to the center of rotation or revolution, and  $\omega$  is the instantaneous angular velocity. For example, the earth in its orbit about the sun has an (average) angular velocity  $\omega$  of  $2 \times 10^{-7}$  rad/sec, and its angular momentum,  $I\omega$ , with respect to the sun is given by

$$\begin{aligned} I\omega &= (m_{\text{earth}} \times r_{\text{earth-sun}}^2) \times \omega \\ &= (5.98 \times 10^{27} \text{ gm}) \times (1.495 \times 10^{13} \text{ cm})^2 \times 2 \times 10^{-7} \text{ rad/sec} \\ &= 2.7 \times 10^{47} \text{ gm-cm}^2/\text{sec}. \end{aligned}$$

Or again, the problem under this heading might involve a flywheel (essentially a metal ring or hoop) of mass  $m$  and radius  $r$  which is to start from rest to reach a certain angular velocity  $\omega$  within  $t$  seconds; what should be the force  $F_1$  applied by way of the belt around the rim? (Fig. 16-18.) Here the torque is  $F_1r$ , consequently  $(F_1r)t = mr^2\omega - 0$ , which we may solve at once for  $F_1$ .

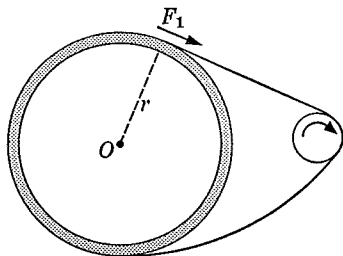


FIG. 16-18.

**PROBLEM 16-18.** The flywheel just mentioned has a diameter of 3 m and a mass of 400 kgm. It is now rotating at 8.3 rad/sec and is to slow down to 0.6 rad/sec in 20 sec by the application of a braking torque along the rim of the wheel. What is the change in angular momentum? How large an (average) braking force must act?

**PROBLEM 16-19.** Find the angular momentum for the earth's spin about its own axis.

**16-9 The law of conservation of angular momentum.** By far the most impressive use of the concept "angular momentum" is to be found in the extension of the law of conservation of momentum to the case of rotation. We may state it in first form as follows:

*In systems on which no torque acts from the outside, the total angular momentum remains constant. Here, too, some concrete example will establish the point best.*

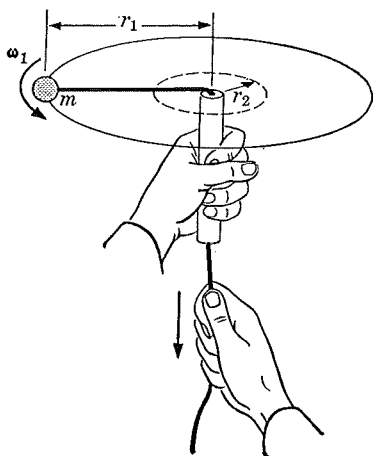


FIG. 16-19. When the right hand pulls the string down, the stone moves in a smaller orbit at a higher speed.

*Example 9.* A small stone is fastened to a light string which passes through a hollow, hand-held tube (Fig. 16-19). The stone is set into rotation so that its angular momentum is  $I_1\omega_1 (= mr^2\omega_1)$ . Now the other hand pulls on the string, thereby shortening the radius of the path from  $r_1$  to  $r_2$ . As this is being done, the observed speed of rotation  $\omega$  increases very markedly from  $\omega_1$  to  $\omega_2$ . The explanation is that

the whole system is a closed one as far as torques are concerned, for the downward pull is directed along the string itself and thus is incapable of providing a torque on the stone. Therefore, by Eq. (16-11),  $\mathfrak{T}$  being zero,  $\Delta I\omega = 0$  (which is the mathematical statement for the law of conservation of angular momentum), or  $I_1\omega_1$  (at the start) =  $I_2\omega_2$  (at the end). We thus write

$$\eta r_1^2 \omega_1 = \eta r_2^2 \omega_2 = \text{constant.}$$

Since  $r_2$  is smaller than  $r_1$ ,  $\omega_2$  is greater than  $\omega_1$  by

$$\omega_2 = \left( \frac{r_1^2}{r_2^2} \right) \omega_1.$$

In short, the angular momentum  $I\omega$  of the stone has been conserved throughout the change in the geometry of the arrangement, the decrease in  $I$  being compensated by a corresponding increase in  $\omega$ . Or, since  $\omega = v/r$ , we might have preferred to write

$$I_1\omega_1 = I_2\omega_2,$$

$$mr_1^2 \frac{v_1}{r_1} = mr_2^2 \frac{v_2}{r_2} \quad \text{or} \quad mr_1 v_1 = mr_2 v_2,$$

therefore

$$v_2 = v_1 \frac{r_1}{r_2}.$$

**PROBLEM 16-20.** Explain in like manner the following observations as manifestations of the law of conservation of angular momentum (to be abbreviated LCAMm): (a) When a stopper is pulled from a filled basin and the water drains out while circulating like a miniature whirlpool, the rate of circulation is largest near the orifice. (b) The destructive effect of a tornado or twister is larger near the center of motion than near the edge of the disturbance. (c) A skater begins a pirouette with outstretched arms, shifts all his weight on one skate, draws his arms very close to his body, and while doing so spins faster and faster.

**PROBLEM 16-21.** It will be remembered from Chapter 11 that Newton's theory of gravitation found initially only a very limited acceptance as against the older, established theory of planetary motion proposed by René Descartes. The latter had proposed a mechanical though largely qualitative model whereby it was explained that the planets were caught in and dragged along by a whirlpool or *vortex* of ether particles centered around the sun. In the second edition of the *Principia*, in 1713, Newton felt it necessary to add a few sentences to show that the vortex scheme is not only physically inconsistent, but in particular leads to predictions that quantitatively contradict the actual observations on planetary motion. Now prove, by means of the LCAMm, that as Descartes' hypothetical ether particles would approach the center of the vortex (at the sun), their orbital speeds  $v_1$  and  $v_2$  at different

distances  $R_1$  and  $R_2$  (Fig. 16-20) should be related according to the expression  $v_1/v_2 = R_2/R_1$ . Since  $v = 2\pi R/T$ , prove next that in this model the periods of revolution  $T$  should be related to the distance  $R$  from the sun by the expression  $T \propto R^2$ . And lastly show that this derived result is in contradiction to the actually observed relationship between  $T$  and  $R$  as expressed by Kepler's Third Law.

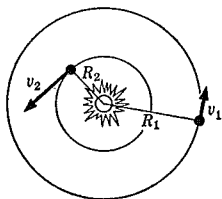


FIG. 16-20.

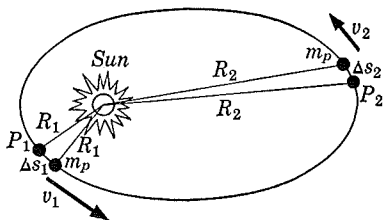


FIG. 16-21.

*Example 10.* Now follows one of the most striking applications of the LCAMm. Let us adopt Newton's fundamental assumption that there are no external forces acting on our solar system, that the only force experienced by each planet is a central force directed toward the sun (except for the small perturbations owing to relatively feeble mutual attractions among the planets themselves). This assumption at once permits us to apply the LCAMm to the elliptical motion of each planet around the sun. For if we reckon all distances  $R$  to the center of force, the sun at the focus of the ellipse, then  $I\omega$  of the planet, namely,  $m_p R^2 \omega$  or  $m_p Rv$ , must be constant at all points along the path, even though the distances and speeds change from point to point (Fig. 16-21). For any two portions of the orbit,

$$m_p R_1 v_1 = m_p R_2 v_2.$$

But clearly the speed  $v$  corresponds to the ratio of some small distance  $\Delta s$  along the orbit to the time  $t$  taken by that distance, or

$$v = \frac{\Delta s}{t} \quad \therefore m_p R_1 \left( \frac{\Delta s_1}{t_1} \right) = m_p R_2 \left( \frac{\Delta s_2}{t_2} \right).$$

If we direct our attention at the planet near positions  $P_1$  and  $P_2$  for equal time intervals,  $t_1 = t_2$  and

$$R_1 \Delta s_1 = R_2 \Delta s_2.$$

Multiplying by  $\frac{1}{2}$  on both sides, we obtain

$$\frac{1}{2} R_1 \Delta s_1 = \frac{1}{2} R_2 \Delta s_2. \tag{16-12}$$

But this is precisely the same information as Eq. (9-1), which was the mathematical statement of Kepler's Second Law, for  $\frac{1}{2} R \Delta s$  is nothing but the area swept out by the radial line  $R$  from the sun to the planet while it covers distance  $\Delta s$ . Therefore Eq. (16-12), which here we have *derived* theoretically from the LCAMm in a rather more elegant and direct manner than by the geometrical argument in Section 9-3, states simply that during equal time intervals  $t$  a planet moving in an elliptical orbit under central forces must follow an *equal area law*. And conversely, now that the prediction implied in Eq. (16-12) is found to coincide with well-known empirical fact, we may place great faith in the correctness of the underlying assumption, namely, the assumption that planets are by and large acted upon only by a central force toward the sun. We might say that the motion of the planets furnishes visible proof of the reign of the law of conservation of angular momentum.

**PROBLEM 16-22.** In the hydrogen atom we may imagine that there is a small positively charged nucleus around which a single negative electron travels in circular or elliptical orbits. The mass of the electron is about  $10^{-27}$  gm, and its radius in circular motion is of the order of  $10^{-8}$  cm. If at that distance the electric attraction which provides the centripetal force on the electron is about  $25 \times 10^{-4}$  dyne, what is its angular momentum? If Kepler's Laws apply also to the motion of such an electron (as they do with proper changes of reference, e.g., substitution of "nucleus" for "sun"), then calculate the angular momentum for the same electron at a radius four times larger than before.

**16-10 Conclusion.** It is possible to extend this discussion of momentum, both linear and angular, to demonstrate more fully the far-reaching power of these concepts. In particular, it is perhaps a little startling to realize that, like linear momentum  $mv$ , the angular momentum of a body is a *vector* quantity. Consequently a change of axis of rotation of spinning disks, spheres, etc., corresponds to a change of angular momentum, just as a change in direction of motion itself implies a change in  $\mathfrak{M}$ . The LCAMm, if interpreted as a vector proposition, can then explain why the plane of rotation of spinning projectiles, gyroscopes, etc., will be maintained in the absence of externally applied torques, and why even a large applied torque may change the plane of spin only slightly.

However, the main point has been presented. The conservation laws of momentum are enormously versatile tools for the analysis of physical behavior of systems, even when the details of motion or

forces are not accessible to direct study. We were able, for instance, by regarding the solar system as a closed system, to derive from the conservation law the Second Law of Kepler without having included any assumption either of the physical nature of the mutual central forces or of the particular paths followed. Consequently, we find that in practice the physical scientist attacks actual problems through the application of the conservation laws of momentum (along with that concerning energy, now to be discussed) more frequently than perhaps by any other method. Therefore we shall not be surprised to meet frequent references to the LCMm and LCAMm in the following chapters.

### Further Reading

P. Lenard, *op. cit.* (Ch. 9); biographies of Descartes, Huygens, Newton, Leibniz.

E. Mach, *The Science of Mechanics*. Chicago: Open Court Publishing Co. (1942, original edition 1883), particularly pp. 396-418. The mathematical arguments may be largely skipped at this point.

W. F. Magie, *op. cit.* (Ch. 2), pp. 50-51. Excerpt from Descartes.

## CHAPTER 17

### THE LAW OF CONSERVATION OF ENERGY

**17-1 Vis viva.** The so-called fundamental quantities of mechanics are usually taken to be length, time, and mass. So far, we have used combinations of these three to set up derived concepts: velocity, acceleration, force, momentum, rotational inertia. And each of these has found a prominent role in some fundamental law of nature. Now we are about to introduce the last of the great mechanical concepts, *energy*, together with the law over which it reigns, the mighty principle of conservation of energy.

The full historical development of the concept and the law was lengthy, and more involved than usual; it took over 150 years from its first adumbration to the point where even the basic terminology itself was fully worked out.\* Yet the roots lie in the very same



Fig. 17-1. Christian Huygens (1629-1695).

---

\* For one of the more recent analyses, see Chapter XXIII in W. P. D. Wightman's *The Growth of Scientific Ideas*, New Haven: Yale University Press (1951).

problem that gave rise to the concept of momentum and the law of conservation of momentum, namely, the problem of how the motion of bodies changes when they collide with one another. And significantly, in introducing a first major contribution, we present the same man who over 30 years before had helped formulate the law of conservation of momentum in elastic impact: Christian Huygens (1629-1695), the Dutch physicist of whom it has often been said that he was in many respects Galileo's and Newton's peer.

Huygens was not only an outstanding mathematician (Newton referred to him as one of the three "greatest geometers of our time"), he also made an improved telescope, discovered a satellite of Saturn, and, building on Galileo's work, invented the first practical pendulum clock. Among his great accomplishments in physics are the first theorems of centripetal force; the conservation laws in elastic impact; the theory of oscillating systems; and a treatise which laid the foundations for the wave theory of light.

The starting point for our topic is Huygens' observation that during the collision of two perfectly elastic spheres ". . . the sum of the product of their masses times the square of their velocities will remain the same before and after the impact . . ." (From the posthumous treatise *De motu corporum ex percussione*, 1703.)

We may discuss this postulate in terms of a concrete but general example, as in Fig. 17-2. Two unequal but perfectly rigid spheres approach each other with unequal speed, collide, and separate. To be sure, from our previous discussion we know well that the law of conservation of momentum demands

$$m_A \mathbf{v}_{A_1} + m_B \mathbf{v}_{B_1} = m_A \mathbf{v}_{A_2} + m_B \mathbf{v}_{B_2}, \quad (17-1)$$

but this is not the issue any longer. What Huygens is proposing here is that there is fulfilled at the same time *another* equation, namely,

$$m_A v_{A_1}^2 + m_B v_{B_1}^2 = m_A v_{A_2}^2 + m_B v_{B_2}^2. \quad (17-2)$$

To the quantity  $mv^2$ , the German philosopher and scientist Gott-

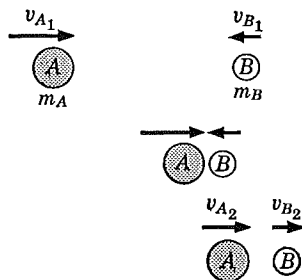


FIG. 17-2. Collision between two moving spheres.



fried Wilhelm Leibniz had previously (1695) given the name *vis viva*,\* so Huygens' statement, and its form in Eq. (17-2), may be called the law of conservation of *vis viva* during perfectly elastic impact.

This is really a very astonishing proposal! Let there be no mistake about it, Eq. (17-1) does not by itself lead to Eq. (17-2). In the latter case the squares of the speeds are involved, not the velocity vectors themselves. Also, Eq. (17-2) is a scalar equation; that is, the directions of motion do not enter at all. Consequently, it holds true in this form even if body *A* were to overtake body *B*, or if it be overtaken by *B*, or if *A* and *B* meet at any angle whatever. A couple of examples will illustrate the usefulness of the new law.

*Example 1.* Recall Example 2 in Chapter 16. A ball *A* of mass  $m_A$  was shot at velocity  $v_{A_1}$  toward a ball *B* of equal mass ( $m_B = m_A$ ) resting on a table ( $v_{B_1} = 0$ ). We inquired:  $v_{B_2} = ?$  With the LCMm alone, this could not be solved unless we added to the four data a fifth, namely,  $v_{A_2} = 0$ , the reason being that the LCMm alone gives only one equation [Eq. (16-3)] with six quantities:

$$m_A v_{A_1} + m_B v_{B_1} = m_A v_{A_2} + m_B v_{B_2}. \quad (16-3)$$

But if we also call upon Eq. (17-2),

$$m_A v_{A_1}^2 + m_B v_{B_1}^2 = m_A v_{A_2}^2 + m_B v_{B_2}^2, \quad (17-2)$$

we have two equations in six quantities, therefore we need only four data to solve for any two unknowns. For example, using what is given above concerning the two masses and the initial velocities, we may calculate directly the values of  $v_{A_2}$  and  $v_{B_2}$ :

$$\left. \begin{array}{l} \text{From Eq. (16-3):} \quad v_{A_1} + 0 = v_{A_2} + v_{B_2} \\ \text{From Eq. (17-2):} \quad v_{A_1}^2 + 0 = v_{A_2}^2 + v_{B_2}^2 \end{array} \right\} \quad (17-3)$$

Squaring the first and subtracting, yields

$$0 = 2 v_{A_2} v_{B_2}.$$

Therefore either  $v_{A_2}$  or  $v_{B_2}$  is zero, and the other one of the two final velocities, by Eq. (17-3), is equal to  $v_{A_1}$ . But if ball *B* is solid,  $v_{B_2}$  cannot be zero while ball *A* passes forward with  $v_{A_2} = v_{A_1}$ . Consequently, the other alternative is the correct one, namely,  $v_{A_2} = 0$ ,  $v_{B_2} = v_{A_1}$ . Now this is a much more satisfactory solution, because we can calculate *both* final velocities exclusively from data that refer only to the initial conditions before impact.

\* *Vis viva* = living force. This proved to be an unfortunate term. Leibniz thought he was defining a quantity analogous to Newton's force (i.e.,  $ma$  or  $\Delta \mathfrak{M}/t$ ), but this was mistaken: eventually the product  $mv^2$  was associated with another name, *energy*. Of this, more in the sequel.

This is true in general for any problem involving direct collision of two perfectly elastic objects. With the aid of the LCMm and Eq. (17-2), the final velocities of both bodies can be found from the initial conditions without further assumptions or data.

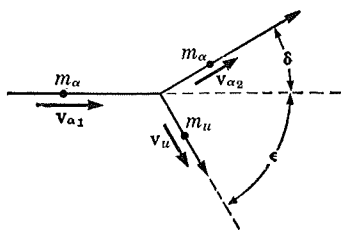


FIG. 17-3.

*Example 2.* An alpha-particle, mass  $m_\alpha$ , is shot at known speed  $v_{\alpha 1}$  into a cloud chamber, makes an elastic but glancing collision with a nucleus of unknown mass  $m_u$ , and is seen to go off at an angle of  $90^\circ$  with respect to the recoil particle (see Fig. 17-3). The known or directly measurable quantities are  $m_\alpha$ ,  $v_{\alpha 1}$  and the angles  $\delta$  and  $\epsilon$ ; also that  $\delta + \epsilon = 90^\circ$  in this example. The unknown quantities are:  $m_u = ?$   $v_u = ?$   $v_{\alpha 2} = ?$  This is a typical problem from nuclear physics, and the solution would proceed as follows:

Conservation of *vis viva* during elastic collision requires that

$$m_\alpha v_{\alpha 1}^2 = m_\alpha v_{\alpha 2}^2 + m_u v_u^2.$$

Conservation of momentum requires that

$$m_\alpha v_{\alpha 1} = m_\alpha v_{\alpha 2} \cos \delta + m_u v_u \cos \epsilon,$$

$$0 = m_\alpha v_{\alpha 2} \sin \delta + m_u v_u \sin \epsilon.$$

From the last two equations, and from the knowledge that here  $\cos \epsilon = \sin \delta$ ,  $\sin \epsilon = \cos \delta$ , it follows that

$$v_{\alpha 2} = -\frac{m_u}{m_\alpha} \frac{v_u}{\tan \delta}, \quad v_{\alpha 1} = -\frac{m_u}{m_\alpha} \frac{v_u}{\sin \delta} \times \cos^2 \delta + \frac{m_u}{m_\alpha} v_u \times \sin \delta.$$

If we substitute these values into the first equation and cancel, we find at once that

$$m_\alpha = m_u.$$

In words: if after an elastic collision the particles are seen to separate along paths at  $90^\circ$  to each other, their masses are equal! In this particular case, therefore, the alpha-particle must have collided with a helium nucleus. In like manner, other recoil particles may be identified.

**PROBLEM 17-1.** Follow the above example further and show that here  $v_{\alpha 2} = v_{\alpha 1} \sqrt{1 + \tan^2 \delta}$ . Then find the expression for  $v_u$  in terms of  $v_{\alpha 1}$  and  $\delta$ .

**PROBLEM 17-2.** Consider again problem 16-12, which had to be solved by means of the LCMm alone. How does the new principle aid such investigation?

PROBLEM 17-3. In the posthumous edition of Huygens' works (1703) we find this "Axiom": "If two equal bodies [bodies of equal mass] moving in opposite directions with equal velocities meet each other directly [in perfectly elastic collision] each of them rebounds with the same velocity with which it came." Show that this axiom is in accord (a) with the LCMm, (b) with the principle of conservation of *vis viva*.

The usefulness of the principle of conservation of *vis viva*, as demonstrated, is considerable, but *unlike* the LCMm *it is strictly limited to impacts between perfectly elastic bodies*. In truth, the matter is even more serious, for it is in general not possible to judge before the collision takes place whether it will be sufficiently close to a perfectly elastic one to admit using the principle; and so we often decide whether the principle applies only after a tentative calculation has shown whether the theoretical results are reasonable. This *a posteriori* procedure is rather like diagnosing whether a sick man has malaria by seeing whether a malaria cure will put him on his feet.

Evidently what is needed is an extension of the principle of conservation of *vis viva* in some manner so that the extended principle will apply to all types of interactions, just as the LCMm applies to all interactions in a closed system. This was brilliantly accomplished in the 19th century and, leaving the main details of the historical development for later, we turn first to an account of the results.

**17-2 Preliminary questions. The pile driver.** Consider a body in motion, perhaps a block sliding on a horizontal plane. If you tried to stop it quickly by interposing your hand in its path, you would feel that the block exerts a force on you, perhaps a considerable one. Is this some new force we have not yet discussed? Do bodies exert force upon obstacles *by virtue of being in motion*, as one might say springs exhibit forces by virtue of being compressed, and magnets by virtue of having inherent magnetism?

First of all, note that this common-sensical and picturesque way of speech, which mirrors the phraseology of the seventeenth and early eighteenth century, explains little and can lead to serious semantic confusions. In the last analysis, as Newton was aware, it only serves to explain magnetism by magnetism, elasticity by elasticity, and so forth. Therefore instead of speaking of an inherent force, we should rather seek the relationship between the force under examination and its observable effects. If we look in this light at the force which moving objects exert on an obstacle like our hand, we see at once

that this is not some new "living force," as was sometimes thought, but rather that our hand experiences the force of reaction to the decelerating or braking force which, by Newton's Second Law, it must apply to the moving object to change its state of motion. This force of deceleration, and also the reaction to it, are of course both numerically given by the product of the mass  $m$  of the object and the deceleration  $a$ . And that is strictly all there is to the original problem.

But it is very revealing to embellish this theme. For example, it has always been of interest to measure with some accuracy the speed of a moving object. Can this not be done by measuring the force needed to stop it? Galileo put the problem this way: consider a falling body; after it has gathered some speed, let it fall upon a vertical stick or stake lightly impaled in some yielding sand (an arrangement that corresponds to the so-called pile driver).

"It is clear," said Galileo, "that if one lifts this body a cubit or two and allows it to fall upon the same material, it will, with this impulse, exert a new and greater pressure than that caused by its mere weight; and this effect is brought about by the [weight of the] falling body together with the velocity acquired during the fall, an effect which will be greater and greater according to the height of the fall, that is, according as the velocity of the falling body becomes greater. From the quality and intensity of the blow we are thus enabled accurately to estimate the speed of a falling body. But tell me, gentlemen, is it not true that if a block be allowed to fall upon a stake from a height of four cubits and drives it into the earth, say, four finger-breadths, that coming from a height of two cubits it will drive the stake a much less distance, and from the height of one cubit a still less distance; . . ." (*Dialogues Concerning Two New Sciences*, Third Day.)

Indeed, are we not thus enabled accurately to estimate the speed of the falling body? Not so simply. By stopping the body, we change its momentum by  $mv$  (numerically). From Newton's Laws

$$mv = F_1 t, \quad \text{and} \quad v = \frac{F_1 t}{m},$$

where  $F_1$  is the (average) net force which decelerates the body (and numerically also gives the force of reaction that in turn pushes on the stake),  $t$  is the time during which the stake is in motion, and  $m$  is the mass of the falling body. Unless we know the factor  $t$  (hardly observable with ease) the force  $F_1$  on the stake gives little information

as to  $v$ . If the sand is tightly packed,  $t$  will be very small,  $F_1$  very large; if the soil is loose,  $t$  is greater, and  $F_1$  is smaller for the same value of  $v$  and  $m$ .

But now, what can we learn about  $v$  from the distance  $s$  through which the stake moves while decelerating the pile driver

from  $v_1$  to  $v_2 = 0$  (Fig. 17-4)? Let us assume that the retarding force  $F_1$  on the falling mass was constant during the collision. Then, since

$$F_1 = ma,$$

$$F_1 \cdot s = mas.$$

But, from Eq. (1-5),

$$as = \frac{v_2^2 - v_1^2}{2},$$

therefore

$$F_1 \cdot s = \frac{1}{2} mv_2^2 - \frac{1}{2} mv_1^2.$$

Since  $v_2 = 0$ ,

$$F_1 \cdot s = \frac{1}{2} mv_1^2 \text{ (numerically).}$$

It appears now that for a given mass  $m$ , and for an obstacle that offers always the same retarding force (a condition nearly enough fulfilled by always using the same stake in the same soil), the distance through which the stake is moved is proportional to the *vis viva*, that is, to the *square* of the speed of the moving body! Using Galileo's example, and assuming that 1 cubit  $\doteq$  50 cm and 1 fingerbreadth  $\doteq$  2 cm, we should rephrase it thus: If a block be allowed to fall upon a stake from a height of 200 cm (with a final velocity  $v$  of  $\sqrt{2gh} \doteq$  620 cm/sec), and drives the stake into the earth a distance  $s$  of, say, 8 cm, the same mass coming from a height of 100 cm (with a final velocity  $v'$  of 440 cm/sec) will drive the same stake a smaller distance  $s'$ , namely (in the ideal case),

$$\frac{s'}{s} = (v'/v)^2, \quad \text{or} \quad s' = 8 \times \frac{440^2}{620^2} \text{ cm} = 4 \text{ cm.}$$

It is only by such a calculation, and under such assumptions, that Galileo's pile driver permits a comparison of relative speeds.

**17-3 The concept of work.** The previous example, to which we shall return, has been chosen for several reasons, not the least being that it introduces, in the historical context, two concepts: *work* and *energy*. The definition of the first of these follows:

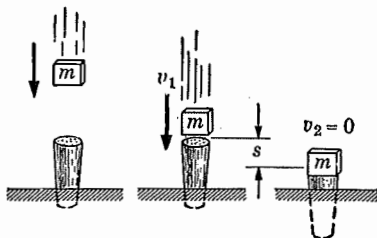


FIG. 17-4. Pile driver.

When a force  $F_1$  acts on a body during a displacement  $s$ , then we say that *the product  $F_1 \times s \times \cos \theta$  is the work done by  $F_1$  on the body*, where  $\theta$  is the angle between the directions of the two vectors  $F_1$  and  $s$ . Thus in the previous discussion, the work done by the force  $F_1$ , the action of the pile driver on the stake, was  $F_1 \times s$ ,  $\cos \theta$  being unity in that case (why?). At the same time, the work done by the stake on the falling mass was  $F_1 \times s \times (-1)$  or  $-F_1 \times s$  (because the displacement  $s$  and the force  $F_1$  were oppositely directed, and  $\cos 180^\circ = -1$ ). Or again, when a horse pulls on a cart as shown in Fig. 17-5, exerting a pull  $F_1$  of 100 lbf (or about  $4.5 \times 10^7$  dynes) at

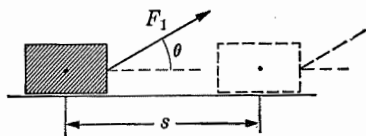


FIG. 17-5. The work done by force  $F_1$  on the block is  $F_1 s \cos \theta$ .

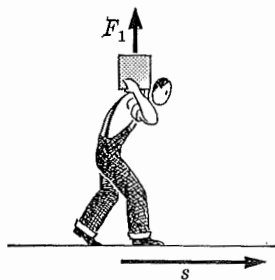


FIG. 17-6.

an angle  $\theta$  of  $30^\circ$  over a distance of 1 ft (or about 30 cm), the work done by the horse on the cart is  $100 \text{ lbf} \times 1 \text{ ft} \times 0.87 = 87 \text{ ft-lbf}$  (or  $4.5 \times 10^7 \text{ dynes} \times 30 \text{ cm} \times 0.87 = 1.17 \times 10^9 \text{ dynes-cm}$ ). On the other hand, if a man carries a load on his shoulders over a distance  $s$ , and supplies to it a balancing force  $F_1$  (upward), the work he does is zero; for here  $F_1$  and  $s$  may both be large, but  $\cos 90^\circ = 0$  (Fig. 17-6). Again, the centripetal force  $F_c$  on an object in circular motion does no work either, for it is directed at  $90^\circ$  to the path at every instant. Lastly, a hard push on a *steady* object may mean  $F_1$  is large,  $\cos \theta = 1$ , but  $s = 0$ ; and although it might tire us out, we have not thereby done *work* as defined in physical science. Evidently, we must not let our subjective notions of activity, achievement, or fatigue become confused with this very different use of the word "work."

Before we go into the usefulness of the concept, note that work is a *scalar* quantity, although it is a product of two vector quantities. The units of work are those of force times distance, i.e., lbf-ft (or rather ft-lbf to distinguish them from those of the very different concept, *torque*) and, in the cgs system, dyne-cm, for which the name *erg* is used. *One erg* being a rather small quantity of work, it is

convenient to refer to a larger unit; by definition,  $10^7$  ergs  $\equiv$  1 joule, named so in honor of the British scientist whose contributions will figure prominently in the following pages.

**PROBLEM 17-4.** A boy pulls a sled with a force of 10 lbf *via* a horizontal string over a distance of 1 mile on level ground. (a) What work does the boy do on the sled, in ft-lbf? in ergs? in joules? (b) How many ergs are there per ft-lbf? (c) If the sled moves at a constant speed, what is the work done by the force of friction upon the sled? (d) What is the work done upon the sled by the gravitational force which the earth exerts on it? (e) If the net work done on an object is simply the sum of the work done by each individual acting force, what is the net work on this object? Show that this net work is the same as the work done on it by the *net force*.

**PROBLEM 17-5.** In the example involving the pile driver, Fig. 17-4, assume that the distance  $s$  was 8 cm and the average force needed to stop the falling body  $10^6$  dynes. (a) Find the work done by the pile driver on the stake. (b) Find the work done by the stake on the driver. (c) If the driver fell from a height of 400 cm and had a mass of  $2 \times 10^4$  gm, what work was done upon it during the descent before collision by the gravitational force?

**PROBLEM 17-6.** A block is pulled without acceleration across a plane surface for a distance of 10 m, as shown in Fig. 17-5. Its mass is 1.6 kgm, the coefficient of friction  $\mu = 0.2$ , the angle  $\theta = 20^\circ$ . What work is being done on the block by the applied force  $F_1$ ? (Use caution in solving for the force of friction!)

**17-4 Various forms of energy.** Work can be done on a body in many ways, and with many results. The stake did work on the falling pile driver, thereby bringing it to rest; conversely, the driver did work on the stake and so drove it into the ground against the forces of friction. Work is done by the hand in lifting the driver to a high position before the descent, thereby setting up the condition needed for gravity to do work on the falling object and so to let it gather speed. When a spring is compressed or extended between two hands, work is done on it. When a flywheel is started or slowed down, forces act over certain distances, and again work is done. Evidently, we should investigate what happens here systematically and quantitatively, still by analyzing simple mechanical problems first, before we are ready to handle the full power of these concepts.

(a) *Work done to overcome frictional resistance.* By far the largest expenditure of work on this earth would seem to be work done "against friction." We pull on a block horizontally with an applied force  $F_{ap}$  to move it through distance  $s$  on a level table. If the force is just enough to counteract the force of friction  $f$  and no net force is

left over for acceleration of the block, then the work done by  $F_{ap}$ , namely  $F_{ap} \cdot s$ , is numerically just equal to  $f \cdot s$ . If we now remove  $F_{ap}$ , the block will just sit on the table. The only reward for the work we have done on the block is the transfer of the block from one spot to another, and a little heating up where the surfaces of contact rubbed against each other. If we call this "wasted" or "dissipated" work, it is so only by comparison with other circumstances where it will turn out that some or all of the work done can, on removal of  $F_{ap}$ , be recovered directly.

(b) *Work done to overcome the inertia of bodies.* We concentrate next on a case where friction is absent, e.g., on a block on a friction-free horizontal plane, in equilibrium at some speed  $v_1$  (which might be zero). Now we apply horizontally a force  $F_{ap}$  through distance  $s$  to accelerate the block to speed  $v_2$ . The work done thereby,  $F_{ap} \cdot s$ , is now

$$F_{ap} \times s = mas = m \left( \frac{v_2^2 - v_1^2}{2} \right)$$

$$\text{or} \quad F_{ap} \times s = \frac{1}{2} mv_2^2 - \frac{1}{2} mv_1^2. \quad (17-4)$$

It appears that the work done on it serves to change the *vis viva* of the object or, more properly, a quantity that is just half the *vis viva*, and which about a century ago received its current name, *kinetic energy* (to be abbreviated to KE). Putting Eq. (17-4) more generally, we can say that here

$$F_{ap} \cdot s = \text{change in kinetic energy} \equiv \Delta \text{KE}.$$

For example, a cyclotron is to accelerate a nuclear particle, e.g., a deuteron (the nucleus of a heavy isotope of hydrogen, made of one neutron and a positive particle, a proton, of about the same mass; total mass of deuteron =  $3.3 \times 10^{-24}$  gm). The desired final speed is one-tenth of the velocity of light, namely about  $3 \times 10^9$  cm/sec. If it starts substantially from rest, what work must be done on the deuteron?

*Solution:*  $\Delta \text{KE} = \frac{1}{2} mv_2^2 - \frac{1}{2} mv_1^2$ ; here  $v_1 \doteq 0$ ;  $\Delta \text{KE} = \frac{1}{2} \times 3.3 \times 10^{-24} \text{ gm} \times (3 \times 10^9 \text{ cm/sec})^2 = 1.5 \times 10^{-5} \text{ gm-cm}^2/\text{sec}^2 = 1.5 \times 10^{-5} \text{ erg}$ .<sup>\*</sup> By Eq. (17-4), this corresponds also to the work needed (in the absence of other effects such as friction) and this amount of work is actually supplied by letting an electric force act upon the accelerating particle. Even though  $1.5 \times 10^{-5}$  erg may seem little work, the difficulties of supplying it are as varied as the effects obtainable with such high-speed particles!

<sup>\*</sup>  $1 \text{ gm-cm}^2/\text{sec}^2 = 1 \text{ erg}$ , because  $1 \text{ erg} \equiv 1 \text{ dyne-cm} = 1 (\text{gm-cm}/\text{sec}^2) \cdot \text{cm}$ .



Thus, work can be "converted" into energy, and work done exclusively to increase kinetic energy is in a sense directly "recoverable," for when a high-speed object is then allowed to fall upon some target, it can in turn do work on it while giving up its kinetic energy. One may say that energy is the capacity for doing work. The pile driver is a case in point. While the driver of mass  $m$  falls through a distance  $h$ , the constant force of gravity,  $mg$ , does work on it ( $mg \times h$ ), thereby increasing its kinetic energy (from Eq. (17-4),  $mgh = \frac{1}{2}mv_2^2$  if there are no frictional losses on the way down and if  $v_1$ , the starting velocity, is zero). The work done by gravity is, to use graphic terms, "stored up" in the moving object, and becomes available on impact with the stake to drive it into the sand.

Although such pictures are a little too concrete, these are important concepts: work on an object can be converted into kinetic energy (or, later, other forms of energy); conversely the energy of a body may be given up in doing work on something else; and often work is done against friction and may be irrecoverably lost from the mechanical system, but then it does show up as heat.

(c) *Work done to change the level of an object.* Consider this third method of doing work on and imparting energy to a body: simply lift up a heavy object from the floor to a distance  $h$  above it. The constant applied force needed here is just  $mg$ , enough to balance the weight, provided we can neglect air friction and also proceed so slowly that there is sensibly no increase of kinetic energy. If not to friction or to KE, where does the applied work go? It is not lost, for if we let go, the pull of gravity will accelerate the body while it falls through distance  $h$  and so will build up its kinetic energy at the end of the fall to exactly the same number of ergs which we had expended in lifting it. Here we may have recourse to an imaginative fiction and say that the work done in lifting the body was "stored" as *potential energy* (PE) in the general region around the body, in the gravitational *field*, and furthermore, that while the object is then allowed to drop, the potential energy is drawn or given back from that region into the body and converted into kinetic energy. So on raising the level of a mass  $m$  by a distance  $h$ , the work we do goes into potential energy ( $mgh$ ) and, on falling, the "work done by the gravitational force" represents only a withdrawal from a bank account of energy which had been accumulated on the way up, until on returning to the original lower level all of the energy has become kinetic.

One should avoid trying to make too material a picture of this sequence or of the seat of potential energy, and regard it rather as a

heuristic device at this stage, an aid to comprehension without independent truth value. There is really only one big difficulty in the new concept, namely the *level of reference*. If we see a book of mass  $m$  on a desk in the second-floor study and are asked point-blank, What is its potential energy?, it would be of no use to reply  $mgh$ , for there is no obvious way to decide whence to measure  $h$ : from the floor? or from the street below? or from the level of the shop counter where it was originally bought? The truth is that potential energy can always be reckoned with respect to an arbitrary zero or reference level, usually, for convenience, the lowest level to which the body descends in the course of a problematical example. And the reason this can be done is, as you will notice, that we shall always deal with *differences* or *changes* in the potential energy ( $\Delta PE$ ) between two levels. So no harm is done by calling the lowest level  $h = 0$ ; instead, calculations are much more simplified than if we were to reckon all potential energies from some fixed point, say from the center of the earth.

**17-5 The conservation law, first form; and applications.** (a) *The inclined plane:* To summarize the three ways in which a body may dispose of work done upon it by an applied force—work against friction, against inertia, against gravitational forces—consider this very typical problem in mechanics. A block of mass  $m$  approaches the bottom of an inclined plane with a speed  $v_1$  and ascends it for a running distance  $s$  while being pulled by an applied force  $F_{ap}$  as shown in Fig. 17-7. We are also given the angle  $\gamma$  and the coefficient of friction  $\mu$ , and the question is: What part of the work done by  $F_{ap}$  goes toward increasing the kinetic energy of the block? From this one might then find the final speed  $v_2$  at the top of the incline.

*Solution:* The work done by  $F_{ap}$  is  $F_{ap} \cdot s \cdot \cos \theta$  (but here  $\theta$ , the angle between the direction of  $s$  and of  $F_{ap}$ , is zero and  $\cos \theta = 1$ ). This amount is split into three not necessarily equal parts: work done against friction  $f$  is  $f \cdot s$  or  $\mu mg \cos \gamma s$ , since  $f = \mu N = \mu(mg \cos \gamma)$ ; work done to change the kinetic energy ( $\frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2$ ); and work done to change the potential energy,  $mgh$ . Note that it is the level difference  $h$ , or  $s \sin \gamma$ , not the path  $s$ , which counts here. In symbols,

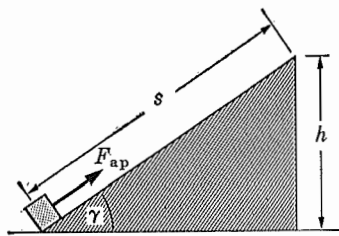


FIG. 17-7. Inclined plane.

$$F_{\text{ap}} \cdot s \cdot (\cos \theta) = \Delta \text{KE} + \Delta \text{PE} + fs, \quad (17-5)$$

$$F_{\text{ap}} \cdot s = \left(\frac{1}{2} mv_2^2 - \frac{1}{2} mv_1^2\right) + (mgh) + \mu mg \cos \gamma \cdot s.$$

Evidently, with  $F_{\text{ap}}$ ,  $s$ ,  $\gamma$ ,  $m$ ,  $\mu$  at hand,  $\Delta \text{KE}$  is quickly found. Note that  $F_{\text{ap}}$  refers to the pull by way of a rope or some other externally applied force, but leaves out of account entirely the force of friction and the force of gravity on the block, for those two are accommodated on the right side of the equation.

Now this was a quite plausible argument, and an experimental check would show that Eq. (17-5) is indeed correct. But is it self-evident, or even proven? Neither! The fact that there are three *obvious* ways in which a body accepts and transforms work done upon it does not mean that the sum of the three terms is necessarily exactly the amount of work done. There could be other drains on the available energy than we have thought of, or perhaps the moving body may receive energy from some source other than the applied force. That this does not happen in simple mechanical phenomena, that Eq. (17-5) is correct, must initially be an experimental discovery, a discovery which amounts to a *law of conservation of energy*; for Eq. (17-5) says that every erg of work supplied to a body can be accounted for by corresponding changes in  $\Delta \text{KE}$ ,  $\Delta \text{PE}$ , and work turned into heat by frictional losses. More generally, imagine a closed system, say a box full of elastic balls, blocks, and inclined planes. Add to the system some energy, perhaps by dropping into the box a marble having a kinetic energy of  $x$  ergs. After a while we look again into the box; it will be found that the bodies by mutual collision have so rearranged their speeds and positions that the energy we added is exactly equal to the sum of  $\Delta \text{PE}$ ,  $\Delta \text{KE}$ , and frictional losses since the moment of interference.

"Dissipative system" is the term applied to a system in which some of the mechanical energy is indeed dissipated irretrievably, as by friction. In contrast, "conservative systems" show no such losses; to them, the law of conservation of energy would apply in simpler form, for if the conservative system were totally isolated, unable to receive energy from or give it up to the rest of the world, then  $\Delta \text{KE} + \Delta \text{PE} = 0$ , a specialized case of the conservation law that was first expressed with most telling power and elegance in Lagrange's *Analytical Mechanics* of 1788, but was used long before in one guise or another.

PROBLEM 17-7. A boy pulls a 3-kgm sled up a  $30^\circ$  slope for a distance  $s = 10$  m (Fig. 17-8). He pulls at an angle  $\theta = 10^\circ$ , starts off at the bottom at a speed of  $\frac{1}{2}$  m/sec, and hopes to get to the top (after approximately con-

stant acceleration) with a speed of 3 m/sec. The coefficient of friction being 0.2, what force must he apply to the sled? How much work does he do on the way up? How much of this work goes into  $\Delta KE$ , into  $\Delta PE$ , and into frictional losses?

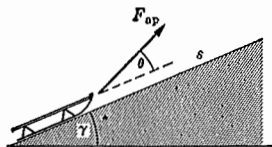


FIG. 17-8.

To carry our argument further, consider a sequel to the problem just given. At the top of the incline, our boy (of mass 20 kgm) sits on the sled and coasts down.  $F_{ap}$  is now zero, and Eq. (17-5) reduces to

$$0 = \Delta KE + \Delta PE + fs$$

or (exchanging sides)

$$\left(\frac{1}{2} m v_{\text{final}}^2 - 0\right) + (-mgh) + \mu mg \cos \gamma \cdot s = 0.$$

Here the change of potential energy is negative: from  $mgh$  at the top to zero at the bottom of the incline. As a *problem*, find the boy's final speed after proving that in this type of motion

$$v_{\text{final}} = \sqrt{2gh(1 - \mu \cot \gamma)}.$$

(b) *Motion where  $F_{ap} = 0$ .* Consider the case of a skidding car which is to be brought to rest by dissipating its kinetic energy through friction with the level road. Here  $F_{ap} = 0$ ,  $h = 0$ ,  $v_2 = 0$ ,  $\gamma = 0$ . What is the distance  $s$  of skidding?

$$(0 - \frac{1}{2} m v_1^2) + (0) + (\mu mgs) = 0,$$

$$\therefore s = \frac{v_1^2}{2\mu g}$$

(if  $v_1 = 60$  mph and  $\mu = 0.25$ ,  $s = 500$  ft!). Note also that the mass  $m$  of the car does not enter into the calculation of  $s$ ; a truck and a motorcycle with equal speeds will skid equal distances.

**PROBLEM 17-8.** A block of mass 2 kgm moves up a  $12^\circ$  incline with an initial speed of  $4 \times 10^3$  cm/sec. Here again,  $F_{ap} = 0$ , so that the mass decelerates uniformly on the way up. From the law of conservation of energy find how far along the incline the object will be when it finally stops and begins to return ( $\mu = 0.2$ ). As in all these strictly mechanical problems, use the conservation law in the form of Eq. (17-5).

**PROBLEM 17-9.** A guillotine knife has a mass of 15 kgm and experiences a frictional force of  $5 \times 10^6$  dynes with the vertical guide rails while traveling up or down. (a) How much work is done by the executioner in raising the blade 3 m from the bottom to the top position? (b) How much kinetic energy will be available when the blade meets the victim?

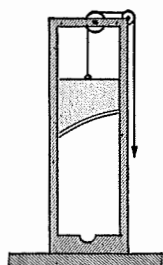


FIG. 17-9.

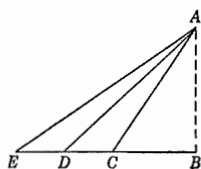


FIG. 17-10.

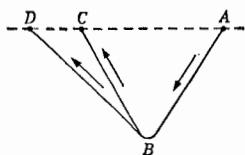


FIG. 17-11. A ball on friction-free planes ascends to its initial level.

**PROBLEM 17-10.** Re-examine Galileo's argument in Section 17-2 again in the light of the law of conservation of energy.

**PROBLEM 17-11.** Two crucial, almost axiomatic propositions on which many of Galileo's arguments in mechanics were based were the following: (a) "The speeds acquired by one and the same body moving down planes of different inclinations are equal when the heights of these planes are equal . . . provided, of course, that there are no chance or outside resistances, and that the planes are hard and smooth." (b) ". . . a body which descends along any inclined [frictionless] plane and continues its motion along a plane inclined upwards will, on account of the momentum acquired, ascend to an equal height above the horizontal, so that if the descent is along  $AB$  the body will be carried up the plane  $BC$  as far as the horizontal line  $ACD$ ; and this is true whether the inclinations of the planes are the same, or are different, as in the case of the planes  $AB$  and  $BD$ ."

Use the Law of Conservation of Energy to prove each of these two propositions separately (Figs. 17-10 and 17-11).

(c) *Pendulums.* Yet another region of application of the law of conservation of energy (to be now abbreviated LCE) relates to the varied and illuminating experiments with pendulums, favorites since the earliest days of modern mechanics. Consider an ideal pendulum, as in Fig. 17-12, released at its topmost position  $a$ , and swinging to the lowest position,  $c$ , before rising again on the other side. While it covers the left half of the arc, from  $a$  to  $c$ , the bob undergoes a vertical displacement

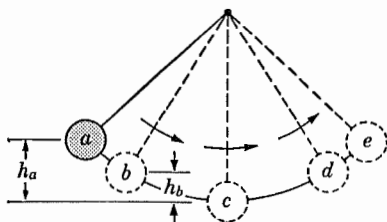


FIG. 17-12. The pendulum.

ment of  $h_a$  cm, and changes its speed from  $v_a = 0$  to  $v_c = \text{maximum}$ . Now we apply the LCE: There is indeed a force  $F_{\text{ap}}$  on the bob (it is the tension in the string) but this acts at  $90^\circ$  to the path of the bob at every instant, and therefore can do no work. If we also neglect friction, the LCE commands here

$$\Delta KE + \Delta PE = 0.$$

For any two points 1 and 2 along the path,

$$\left(\frac{1}{2} mv_2^2 - \frac{1}{2} mv_1^2\right) + (mgh_2 - mgh_1) = 0.$$

Specifically, comparing positions  $a$  and  $c$ , and noting that  $v_a = 0$ ,  $h_c = 0$ ,

$$\left(\frac{1}{2} mv_c^2 - 0\right) + (0 - mgh_a) = 0, \\ \therefore \frac{1}{2} mv_c^2 = mgh_a, \quad \text{or} \quad v_c = \sqrt{2gh_a}.$$

Here it turns out that the speed of the pendulum bob at the lowest point is exactly the same as it would be if the bob were simply to *drop* through the same level difference  $h_a$ . Furthermore, we recognize that the pendulum has to lift itself through the same height on each side of the swing in order to convert all its KE at position  $c$  to PE at  $b$  or  $a$ . Once started, the pendulum must incessantly go through its symmetrical motions if there is no way for its energy to escape, continually exchanging its initial energy back and forth between PE and KE.

**PROBLEM 17-12.** The bob of a simple pendulum (mass = 100 gm) is let go from a vertical height  $h_a$  of 5 cm. (a) What will be its speed at the bottom of the swing? (b) What is its kinetic energy at that point? (c) What is its PE and KE at a point halfway up toward the top position on the other half of the arc?

**PROBLEM 17-13.** One of Galileo's famous arguments involved an arrangement now called Galileo's pendulum. First let a simple pendulum swing freely back and forth between points  $B$  and  $C$  (Fig. 17-13). Now put a nail or a peg in position  $D$ , so that the pendulum, released again from  $B$ , cannot reach  $C$ ; the string is bent at  $D$  and the ball or bob rises to  $E$  before returning. The significant thing is that  $E$  lies on the horizontal line  $BC$ , even though the pendulum on the right half of its trip now swings through the smaller arc  $FE$  rather than  $FC$ . As Galileo has Salviati say in the *Two New Sciences*: "Now, gentlemen, you will observe with pleasure that the ball swings to the point  $E$  on the horizontal, and you would see the same thing happen if the

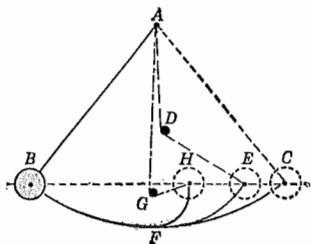


FIG. 17-13.

obstacle were placed at some lower point, say at  $G$ , about which the ball would describe arc  $FH$ , the rise of the ball always terminating exactly on the line  $BC$ ." Explain these observations by means of the LCE.

PROBLEM 17-14. In Huygens' *Horologium Oscillatorium* (1673), we find this proposition: "If a simple pendulum swings with its greatest lateral oscillation, that is, if it descends through the whole quadrant of a circle [i.e., starts, as in Fig. 17-14, from a horizontal position], when it comes to the lowest point in the circumference, it stretches the string with three times as great a force as it would if it were simply suspended by it." Prove this with the aid of the LCE.

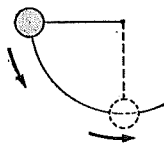


FIG. 17-14.

PROBLEM 17-15. The measurement of the velocity of a fast projectile has always been difficult. One early solution was found in the so-called "ballistic pendulum" of Benjamin Robins (1742). The bullet of known mass  $m_b$  and unknown velocity  $v_b$  is fired horizontally toward and embeds itself in a bob (mass  $m_p$ ) of a freely hanging simple pendulum. The maximum height  $h$  to which the pendulum rises after the inelastic collision with the bullet is noted. (a) With the aid of the LCMm, derive the equation for the initial speed of the pendulum after impact. (b) Derive the expression for  $v_b$  before impact in terms of  $m_b$ ,  $m_p$ ,  $g$ , and  $h$ , using the LCE for analyzing the motion of the pendulum.

(d) *Collisions*. It is surely now evident that the law of conservation of *vis viva*, the original stimulus for the more general LCE, is only a special case of the latter. For if a set of spheres has a perfectly elastic collision, then *by the very definition* of the phrase "perfectly elastic" we mean that none of the kinetic energy of the colliding bodies is lost to frictional processes, i.e., that the system of bodies is a conservative one. For such collisions, either in a horizontal plane or, more generally, of such brief duration that the change in level is insignificant during the time interval, the LCE, which commands that

$$\Delta KE + \Delta PE + \text{work lost to friction} = 0,$$

becomes simply  $\Delta KE = 0$ .

For several colliding bodies, this means

$$\left(\frac{1}{2} m_A v_{A_2}^2 - \frac{1}{2} m_A v_{A_1}^2\right) + \left(\frac{1}{2} m_B v_{B_2}^2 - \frac{1}{2} m_B v_{B_1}^2\right) + \dots = 0.$$

Comparison with Eq. (17-2) shows that this is exactly the same as the law of conservation of *vis viva* (the factor  $\frac{1}{2}$  may, of course, here be canceled throughout). Furthermore, we now have in principle

a way to calculate what happens in all collisions, including those which are not elastic and in which *vis viva* (or kinetic energy alone) is not conserved. While it is, in general, not easy to judge how much energy is lost to friction, the case is not at all hopeless; one type of solution will suggest itself in connection with the following problem.

**PROBLEM 17-16.** A certain ball is generally observed to bounce back from the floor to  $\frac{1}{3}$  of the height of release after free fall. If it has a mass of 5 gm and is thrown vertically downward from a height of 100 m with an initial speed of 50 m/sec, where will the ball be after 2.5 sec? What will be its KE and PE at that moment?

**PROBLEM 17-17.** A widely used demonstration involves a series of equally heavy white balls arranged in one line on a smooth table (Fig. 17-15). Now a black ball (like the others, a perfectly elastic one) is shot against them at speed  $v$ . If its mass is equal to that of each of the others, the last white ball flies off with the same speed  $v$ . But if its mass were twice as large, we would never see *one* white ball fly off at a greater speed, but all white balls move at different speeds. Show that this satisfies both the LCMm and the LCE. If necessary, see Kline, *Am. J. Phys.*, **28**, 102 (1960).

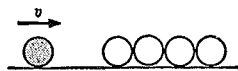


FIG. 17-15.

**17-6 First extension of the conservation law.** For the sake of initial simplicity, we have so far avoided several sets of important problems to which the LCE in its present form would not apply. This does not at all mean that the LCE has only a limited validity! Rather, it means that we must extend the meaning of PE and KE a little to cover some other ways in which energy may appear or be stored up.

(a) *Potential energy of elastic bodies.* Work or KE is transformed into PE when it serves to lift a massive object through some distance. But besides gravitational PE there are other forms of PE, for example, *elastic PE*. If a force acts on a spring (as in Fig. 17-16) to compress it for a distance  $x$ , the work done on the spring is the average force times the displacement. Specifically, the force needed to compress a spring is proportional to the compression, and so it changes linearly from zero at the start to  $F_e = (\text{a constant } \kappa) \times x$  at the end;  $F_{\text{average}} = \frac{1}{2} \kappa x$ , and the work done is  $(\frac{1}{2} \kappa x) \cdot x = \frac{1}{2} \kappa x^2$ . This last term can well be called the potential energy of the spring, for as soon as  $F_e$  is removed, the spring will snap back and make available (to a machine, a clock, or in some other way) all the energy which has been given to it in the initial compression.



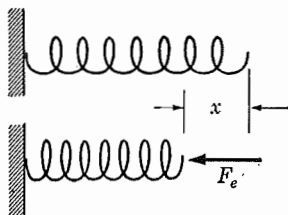


FIG. 17-16.

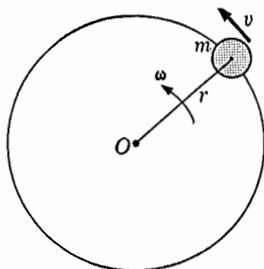


FIG. 17-17.

Of course, the same argument applies to the extension of a spring instead of its compression or, more generally, to the elastic deformation of all bodies. Thus when two elastic spheres collide, their KE is briefly converted to the PE of elastic deformation while they are in contact, only to be completely reconverted to KE as the spheres separate again after impact.

(b) *Kinetic energy of rotation.* A particle of mass  $m$  in rotation about a center  $O$  at angular speed  $\omega$  (Fig. 17-17), has a kinetic energy  $\frac{1}{2}mv^2$ , which we may write also as  $\frac{1}{2}m(r^2\omega^2)$ . But we recognize the factor  $mr^2$ ; it is the rotational inertia  $I$  of the particle. Therefore

$$\text{KE}_{\text{rotation}} = \frac{1}{2}I\omega^2. \quad (17-6)$$

This last equation does not apply only to a rotating particle, but to any rotating object whatever,  $I$  standing for the appropriate rotational inertia in each case,  $\omega$  for the instantaneous angular speed. For example, a disk of radius  $r$  and mass  $m$  rolling at a constant speed on a horizontal plane has a total kinetic energy made up of two parts: translational KE by virtue of the forward motion of the wheel as a whole with speed  $v$  (i.e.,  $\text{KE}_{\text{trans}} = \frac{1}{2}mv^2$ ), and rotational KE owing to the rotation of the rim about the center at angular speed  $\omega = v/r$ , i.e.,  $\text{KE}_{\text{rot}} = \frac{1}{2}I\omega^2 = \frac{1}{2}(\frac{1}{2}mr^2) \times (v^2/r^2) = \frac{1}{4}mv^2$ .

$$\text{KE}_{\text{total}} = \text{KE}_{\text{trans}} + \text{KE}_{\text{rot}}. \quad (17-7)$$

*Example 1.* A disk of mass  $m$  and radius  $r$  begins to roll freely from the top of an incline to the bottom, i.e., to a level of  $h$  cm below. What is its final speed, if rolling friction may be neglected?

*Solution:* We set up the LCE in expanded form:  $F_{\text{ap}} = 0$ ,  $v_1$  and  $\omega_1$  at top = 0,  $f = 0$ . Here,

$$\begin{aligned} \Delta\text{PE} + \Delta\text{KE}_{\text{trans}} + \Delta\text{KE}_{\text{rot}} &= 0, \\ (0 - mgh) + (\frac{1}{2}mv^2 - 0) + (\frac{1}{2}I\omega^2 - 0) &= 0. \end{aligned}$$

But  $\omega$ , the angular speed at the bottom, is given by  $\omega = v/r$ , and  $I_{\text{disk}} = \frac{1}{2} mr^2$ . Substituting and transposing,

$$mgh = \frac{1}{2} mv^2 + \frac{1}{2} \left( \frac{1}{2} mr^2 \right) \frac{v^2}{r^2},$$

$$gh = \frac{1}{2} v^2 + \frac{1}{4} v^2,$$

$$v = \sqrt{\frac{8}{3} gh}.$$

We note, perhaps with initial astonishment, that the final speed  $v$  is independent of mass and radius. All disks should therefore have the same final speed for the same descent, and  $v$  for any rolling disk is less (by  $\sqrt{0.07 gh}$ ) than that for an object sliding down without friction.

**PROBLEM 17-18.** A hoop and a disk of equal radius  $r$  and equal mass  $m$  are released side by side along an incline. Soon the disk has outdistanced the hoop (Fig. 17-18). (a) Calculate and compare the respective final velocities. (b) Where is the hoop at the instant that the disk has reached the bottom?

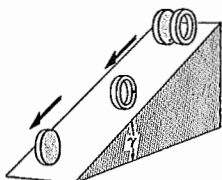


FIG. 17-18.

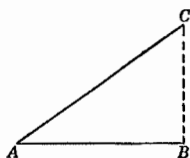


FIG. 17-19.

**PROBLEM 17-19.** Give a physical explanation why a sliding body reaches the bottom of an incline before rolling bodies do.

**PROBLEM 17-20.** Discuss why Galileo's following statement is wrong, and how it should be corrected: "All resistance and opposition having been removed, my reason tells me at once that a heavy and perfectly round ball descending along the lines  $CA$  [and]  $CB$  [a vertical free fall] would reach the terminal points  $A$  [and]  $B$  with equal momenta." (Fig. 17-19.)

**PROBLEM 17-21.** By experimenting with *rolling* spheres on smooth planes Galileo could indeed confirm that the acceleration is constant (see Section 2-4), but not having the concept of rotational inertia, he could not from such data calculate the value of acceleration due to gravity ( $g$ ) on freely falling bodies. Explain why.

**17-7 Further extension and applications of the LCE.** Apart from the elastic potential energy and kinetic energy of rotation, there are a number of other physical processes which can add terms to the energy equation. For example, if some of the bodies are electrically charged, then there is a change in electrical potential energy as they approach or recede from one another, and this has to be included in

$\Delta PE_{\text{total}}$ . Or if chemical energies are set free in a closed system, they must be accounted for. Heat, too, as will be seen at once, is to be counted as a form of energy, so that heat supplied to a system, or generated within by the transformation of other forms of energy to heat, must appear in our equations. So also must sound, and light, and other radiations which carry with them energy.

Altogether, the complete story would seem to produce a very bulky equation for the LCE, something like this:

External work done on, or other energy supplied to, a system of bodies =  $\Sigma[\Delta(\text{PE of all types}) + \Delta(\text{KE of all types}) + \text{frictional losses} + \Delta(\text{chemical energies}) \dots]$ .

Although the complete LCE looks formidable indeed, in actual practice one encounters almost always only such cases as we have seen already, namely, cases in which only two or three of the many types of energy changes are of importance. But every time we make use of it, we are struck by the generality of meaning of the LCE. First of all, energy cannot be created or destroyed without a corresponding opposite change happening somewhere else in the universe, (e.g., the kinetic energy I impart to a cart or a baseball comes ultimately from a diminution of chemical energy within my muscles). Secondly, energy in any one form can be transformed into energy of some other form (a magnet pulling an iron nail toward it shows how magnetic potential energy is changed into mechanical kinetic energy).

(a) *Application to machines.* The LCE can be applied in very simple form to machines, which are all essentially devices for transferring (more or less completely) the energy represented by work done upon one part of the system to some other part of the system, and making it available there in some convenient way. As one example, consider the lever in Fig. 17-20. The system here consists of a straight bar supported on a pivot or *fulcrum* that divides its length in the proportion  $l_1:l_2$ . A load  $F_2$  ( $=mg$ ) on the left is kept balanced by an applied force  $F_1$  on the right. To lift the load, one need increase  $F_1$  only infinitely little, provided the friction at the fulcrum is negligible. While  $F_1$  depresses one end through a small arc  $s_1$ , the other rises through arc  $s_2$ . If the angle covered is very small,  $s_1$  and  $s_2$  are sufficiently close to being straight

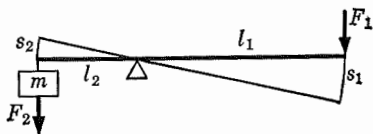


FIG. 17-20. The lever.

lines and perpendicular to the horizontal, so that the work done on the system by  $F_1$  is  $F_1 s_1$ . This work serves to change the KE and PE of load and lever. But if, now, the speed of operation is kept very small,  $\Delta KE = 0$ , and if also the lever is relatively light, the only significant job that  $F_1 s_1$  can do is to increase the PE of the load ( $\Delta PE = F_2 \times s_2 = mgs_2$ ). In the ideal case,

$$F_1 s_1 = F_2 s_2, \quad \text{or} \quad F_1 = F_2 \frac{s_2}{s_1}.$$

By similar triangles from Fig. 7-20, note that  $s_2 : s_1 = l_2 : l_1$ . On substituting,

$$F_1 = F_2 \frac{l_2}{l_1}, \quad \text{or} \quad F_1 : F_2 = l_2 : l_1. \quad (17-8)$$

In words: the force needed to balance or slowly lift a load by means of a simple lever is to the load as the inverse ratio of the respective lever arms  $l_1$  and  $l_2$ . This is the famous law of the lever, known empirically throughout antiquity, and derivable also from other axioms, as is best illustrated by the famous work of Archimedes of Syracuse (287-212 B.C.). The problem was also discussed by Aristotle, Leonardo da Vinci, Galileo, and many others.

**PROBLEM 17-22.** Calculate first how much the earth might weigh, then calculate the dimensions and sketch the position of a lever that would have enabled Archimedes to carry out his thought of lifting the earth from its place in our solar system if only he were given a long enough lever (and a fulcrum).

An important principle from another branch of mechanics is hidden behind Eq. (17-8). If we write it as

$$F_1 l_1 = F_2 l_2,$$

then the meaning is: for equilibrium of the lever, the torque on the right-hand member (tending to produce a clockwise motion) must be balanced by the torque on the left side (which alone would give a counterclockwise motion). More generally, one can say that a body achieves rotational equilibrium, i.e., has no angular acceleration, only if all clockwise torques balance all counterclockwise torques. This is the principle of equilibrium as regards rotational motion; it supplements the principle of equilibrium for translational motion, namely, Newton's First Law, which commands that the forces themselves must balance if a body wishes to avoid linear acceleration. The latter is called the *first principle of statics*; the former is the *second principle of statics*.

**PROBLEM 17-23.** In Fig. 17-21, a weightless bar is acted upon by two equal forces. Show that although the forces cancel and therefore the first principle of statics (principle of translational equilibrium) is obeyed, the second principle of statics is not; that therefore the object will spin with increasing angular velocity. Then suggest where to apply additional forces, and of what magnitude, to fulfill both conditions of equilibrium.

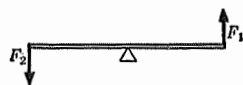


FIG. 17-21.

Like the lever, the pulley is a simple machine well known in ancient times. A more useful and complex version, the simple block-and-tackle (Fig. 17-22), was considered by Aristotle and analyzed by Archimedes. As our hand pulls a distance  $s_1$  on the free end of the string, the lower pulley and the load attached to it are raised through a distance  $s_2$ . Making the same assumptions as in the previous case, the work supplied =  $\Delta PE$  of load, or

$$F_1 s_1 = F_2 s_2, \quad \text{and} \quad F_1 = F_2 \frac{s_2}{s_1}.$$

By measurement, we should find that in this particular pulley system,  $s_2$  is always  $\frac{1}{2} s_1$  (can you see why?), but in more elaborate arrangements the ratio  $s_2/s_1$  may be much smaller. In every case, however, the last equation allows us to predict on the basis of an experimental determination of  $s_2/s_1$  what force is ideally needed to lift a given load, even though we may be entirely mystified by the details of construction and operation of pulley systems!

We may carry this one step further and generalize on all simple mechanical arrangements in the following way. Such machines are equivalent to a black box (Fig. 17-23) to which we supply work

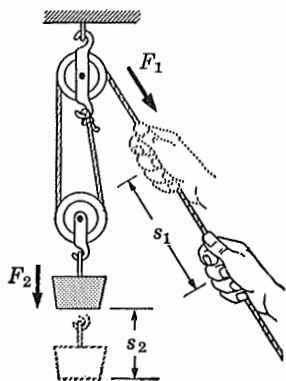


FIG. 17-22. Lifting a load with a block and tackle.

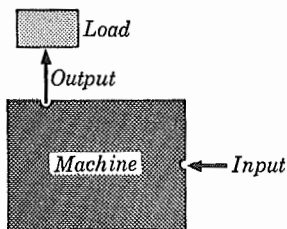


FIG. 17-23. Schematic representation of a machine.

through one opening and get work done on some load by a connection to a second opening. We may know or understand nothing about the gears inside. Yet the LCE demands that the following be always true:

Energy supplied to machine (INPUT)

$$= \text{Energy supplied by machine to load (OUTPUT)} + \text{Energy kept by machine (its own } \Delta KE + \Delta PE + \text{friction losses).}$$

Evidently, a good machine generally is one that does not keep much energy from the load, particularly by losses to friction. If we define the *efficiency* of a machine (symbol  $\eta$ , Greek letter *eta*) as the ratio of *output* to *input*, the ideal machine has  $\eta = 1.00$  (or 100%), although in actual practice we may find  $\eta = 0.1$  (10%) or less.

PROBLEM 17-24. An *ideal*, weightless and frictionless block-and-tackle as shown in Fig. 17-22 would require  $F_1$  to be  $10^6$  dynes if the load is  $2 \times 10^6$  dynes, but in an actual case  $F_1$  may be  $1.3 \times 10^6$  dynes. (a) How much work must be done to raise the same load by one meter? (b) What is the efficiency of this machine in practice? (c) What is the reason for using the machine when it would evidently require so much less energy to lift the load directly?

(b) *Applications to nuclear physics.* (1) *Relativistic effects.* A contribution to physics which in its general implications proved to be most spectacular was contained in a series of papers by Albert Einstein from 1905 on, the first of which bore the unrevealing title "The Electrodynamics of Moving Bodies." It had been known for some time previously that on theoretical grounds the mass or inertia of an electron must increase with increasing speed, so that the acceleration of an already fast-moving electron, by a given force, would be less than that of the same electron at rest. Experimental checks in general harmony with the theory were made by Kaufmann (1902) and others, by deflecting in electric and magnetic fields the high-speed electrons ( $\beta$ -rays) emitted by radioactive nuclei. As we would write it now, Einstein's conclusion was that the "tangible," ponderable mass of any particle, whether charged like an electron or not, increases with speed  $v$  according to the now famous equation

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (17-9)$$

where  $m_0$  is the mass of the particle at rest with respect to the ob-

server, called *rest mass*, and  $m$  is the mass of the particle measured while it moves at speed  $v$  relative to the observer, called *observed* or *relativistic mass*. (This mass determination may be made by supplying a known centripetal force and measuring the radius of curvature of the path, i.e., from  $F = mv^2/r$ .) Furthermore,  $c$  is the velocity of light, a constant, about  $3 \times 10^{10}$  cm/sec.

Now it is evident that for ordinary speed, where  $v$  is much less than  $c$ ,  $m = m_0$  within the experimental error of measurement. But precisely in the motion of atomic and subatomic particles this restriction no longer holds—some  $\beta$ -rays are ejected from disintegrating nuclei with speeds greater than 90% that of light.

At first glance this increase of mass with speed is deeply disturbing. Surely a moving particle has no more molecules, no more “tangible,” ponderable matter than before! If mass increases, mass is not simply a measure of the quantity of matter after all. And then the law of conservation of mass would appear to be no longer strictly true. Such are the speculations that we must consider next.

PROBLEM 17-25. Evaluate the ratio  $m/m_0$  for an object that has successively a speed  $v$  of 0.1, 0.3, 0.5, 0.7, 0.9, and 0.97 times the velocity of light  $c$ . (The experimental check, incidentally, has been found to be very good.) What should be the mass of any object if  $v = c$ ?

PROBLEM 17-26. How much would your mass increase if you were flying in a jet craft at the speed of sound (about 330 m/sec)?

PROBLEM 17-27. As remarked in Chapter 10, the motions of the planet Mercury in its eccentric orbit and of comets that pass near the sun show slight deviations from what is expected on the basis of the Newtonian law of universal gravitation. Attempt a qualitative explanation of these observations in terms of Eq. (17-9). Why is no observable deviation expected for the earth?

The ratio  $(v/c)$  is usually replaced by the symbol  $\beta$ . Thus Eq. (17-9) is written  $m = m_0(1 - \beta^2)^{-\frac{1}{2}}$ . However, the binomial theorem from algebra teaches that  $(1 - \beta^2)^{-\frac{1}{2}} = 1 + \beta^2/2 + 3\beta^4/16 +$  other terms of still higher powers. Now if  $v$  is considerably smaller than  $c$ ,  $\beta^2$  is very small, and  $\beta^4$ , together with all higher terms, is negligible. Therefore we can write

$$m \doteq m_0 \left( 1 + \frac{\beta^2}{2} \right) = m_0 \left( 1 + \frac{v^2}{2c^2} \right) = m_0 + \frac{1}{2} \frac{m_0 v^2}{c^2}.$$

This last result opens our eyes to an amazing physical interpretation of the increase in mass with speed, for it points out directly that the increase in mass is

$$\Delta m \equiv m - m_0 = \frac{\frac{1}{2} m_0 v^2}{c^2} = \frac{\text{KE}}{c^2}.$$

In trying to understand the change in mass with speed, we are led to the concept that kinetic energy added to a ponderable object also adds inertia in an amount  $\text{KE}/c^2$ . Whether we say energy *has* mass or *is* mass or *appears as* mass is only playing with words and adds nothing to the last equation.

Although we have arrived at this result by an approximation, the equation  $\text{KE}/c^2 = \Delta m$  is true in general. Only we must realize that in such a purely mechanical situation *the kinetic energy of a fast-moving particle might be redefined as*

$$\text{KE} = \Delta mc^2,$$

and is not separately evaluated by  $\frac{1}{2} m_0 v^2$ ; for the mass increases from  $m_0$  with speed, and so more kinetic energy is required for equal changes in speed at the latter stages of reaching  $v$  than at the earlier stages.

PROBLEM 17-28. Einstein's original approach was to postulate that no material body can be made to reach a velocity equal to or larger than the speed of light, with respect to the observer. Prove that the foregoing is in harmony with this postulate. (*Hint*: What kinetic energy would be needed if such speeds were desired?)

The idea that it is the increase of energy which changes the mass can now be extended further to include energies other than ordinary kinetic. Consider first a spring with a mass  $m_0$  when limply extended. When we compress it and so give it elastic potential energy PE, its mass increases to  $(m_0 + \text{PE}/c^2)$ . A lump of metal likewise increases its observed mass when it is heated; here  $\Delta m = \text{heat energy supplied}/c^2$ . In brief, the principle of mass-equivalence of energy was soon extended to cover the increase (or decrease) of any type of energy; *for every unit (1 erg) of energy supplied* to a material object, whether it be kinetic energy given to a bullet, gravitational potential energy acquired by the change in relative distance of two stars, or any other form of energy, *the mass of the system thereby increases by*

$\frac{1 \text{ erg}}{(3 \times 10^{10} \text{ cm/sec})^2} = 1.1 \times 10^{-21} \text{ gm}$ . To repeat, this does not mean

there are now in the system more molecules than before; what has changed is the observable inertia of the energy-enriched material. Evidently the rate of exchange, i.e., the factor  $c^2$  in



$$\Delta m = \text{Energy}/c^2,$$

indicates that such changes in mass are beyond direct experience in ordinary mechanical experiments, and therefore there was no need to consider this academic point in the section on the conservation of mass. But in nuclear phenomena or in cyclotrons and other accelerators, where the masses of the particles involved are relatively small to start with and the energies are relatively large, the changes in mass may become very noticeable. If you prefer, *mass* does not measure the amount of tangible or ponderable matter in a system, it measures the amount of matter together with the mechanical energies, electrical energies, etc. Therefore we rewrite the law of conservation of mass to read:

$$\text{In a closed system, } \sum \left( m_0 + \frac{\text{Energy}}{c^2} \right) = \text{constant.} \quad (17-10)$$

A striking illustration of the law in this form is the following example of the complete interconversion between crude, coarse matter and what might seem to be disembodied, radiant energy. At times, very intense gamma-rays (namely, x-rays of short wavelength, and so akin to visible light) may be transformed into material particles, more precisely into pairs of charged particles, one an ordinary negative electron, the other a positively charged electron (a *positron*). Here we may say that before the conversion the mass in the system was (energy of  $\gamma$ -ray/ $c^2$ ), namely, the mass-equivalent of the radiant energy; and after the conversion, the mass in the system is (rest mass of both particles + total KE of particles/ $c^2$ ); the mass so calculated is the same before and after.

Another typical instance of the use and range of these ideas may be found in comparing the mass of a nucleus with the mass of its component particles (nucleons) after separation. For example, a helium nucleus (or  $\alpha$ -particle) has a rest mass of about 4.0028 amu; it is composed of four nucleons, namely, two protons and two neutrons. When the helium nucleus is broken up in a nuclear reaction, the rest mass of each proton measures about 1.0076 amu and that of each neutron about 1.0090 amu. But  $(2 \times 1.0076 + 2 \times 1.0090)$  amu is larger than 4.0028 amu by about 0.03 amu (or, since 1 amu =  $1.66 \times 10^{-24}$  gm, by  $5 \times 10^{-26}$  gm), and it would seem that the sum of the parts "weighs" more than the parent nucleus. And so they do indeed, because energy had to be supplied to break up the helium nucleus, specifically  $5 \times 10^{-26}$  gm

$\times (3 \times 10^{10} \text{ cm/sec})^2 = 4.5 \times 10^{-5} \text{ erg}$ . The complete breakup of helium nuclei is therefore an energy-absorbing (*endothermic*) reaction, and the energy that must be supplied, called the *binding energy* of the nucleus, is stored, as it were, as potential energy among the fragments. Indeed, we shall see that it can be retrieved in the reverse reaction of building up helium nuclei from protons and neutrons.

All that has been said so far demonstrates the *mass-equivalence of energy*. Now nature generally exhibits a symmetry in her operations, that is, one may reasonably suspect that a particle of matter may lose some of its mass by giving up a corresponding amount of energy. This can, of course, be made to occur most directly by slowing down a high-speed particle. It will get "lighter" by an amount given as (loss of  $\text{KE}/c^2$ ) although, of course, whatever agency served to stop the particle must then absorb that much energy and consequently must get that much "heavier" in turn. But there are more striking ways to realize this exchange between mass and energy: The electron and positron mentioned above may again recombine, and then assume once more the character of  $\gamma$ -radiation; ordinary ponderable matter now seems to disappear from view and "pure," "free" radiant energy is observed instead. The balance here is as follows: the relativistic mass  $m$  of the two particles (total rest mass plus mass-equivalent of  $\text{KE}$ ) goes into  $mc^2$  energy units of  $\gamma$ -radiation. Again it is largely a matter of taste whether to say here that matter has been destroyed while energy has been created, or that matter *is* energy or has assumed the *form* of energy, etc. The physical facts behind such phrases are the same, as stated, for example, by Einstein in 1905 as follows:

"If a body gives off the energy  $\Delta E$  in the form of radiation, its mass diminishes by  $\Delta E/c^2$  . . . The mass of a body is a measure of its energy content; if the energy changes by  $\Delta E$ , the mass [of that body] changes in the same sense by  $\Delta E/(3 \times 10^{10})^2$ , the energy being measured in ergs and the mass in grams."

Not always is the change as drastic as in the example of electron-positron "annihilation." In some subatomic phenomena only a small fraction of the mass of the system is transferred into an equivalent quantity of radiant or kinetic energy, as in the release of nuclear energy. But now the factor  $c^2$  is favorable—by the relation  $\Delta E = (\Delta m)c^2$ , any small  $\Delta m$  results in a large  $\Delta E$ .

PROBLEM 17-29. According to the latest figures, the combined installations for producing electric power in the United States have a capacity of about  $10^{17}$  ergs of energy per second. If instead of hydroelectric, steam, and internal combustion engines, one central nuclear power plant were to provide this energy by conversion into electrical energy of the nuclear "fuel," how much matter would be converted per day? How much material would be handled per year, if only 0.08% of the material supplied is converted by this process?

It stands to reason that the law of conservation of energy must be reformulated if energy-mass conversions are to be included. One simple way of doing this is to consider every object in the system as a potential source for complete annihilation, and therefore to assign to every rest mass  $m_0$  a rest energy  $m_0c^2$ , a potential energy which in the course of events may be partly or completely converted into other forms of energy. Then we can safely say:

In a closed system, the total amount of energy, namely, rest energy ( $\Sigma m_0c^2$ ) plus all *other* forms of energy ( $\Sigma E$ ), is constant, or

$$\Sigma(m_0c^2 + E) = \text{constant.} \quad (17-11)$$

Now this is really not a new law; Eq. (17-10), the extended law of conservation of mass, read  $\Sigma(m_0 + E/c^2) = \text{constant}$ , and if Eq. (17-11) is divided through by  $c^2$  on both sides we obtain the very same expression. Either equation tells essentially the same story. As Einstein himself pointed out:

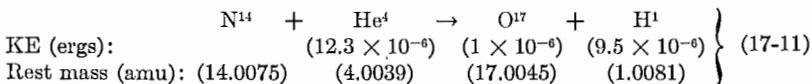
"Pre-relativity physics contains two conservation laws of fundamental importance, namely, the law of conservation of energy and the law of conservation of mass; these two appear there as completely independent of each other. Through relativity theory they melt together into *one* principle."

In this sense one may say that the concepts of mass and of energy also "melt together" as though they were two aspects of one physical reality.

PROBLEM 17-30. If the laws of conservation of mass and of energy are jointly contained either in Eq. (17-10) or Eq. (17-11), how shall we reformulate the law of conservation of linear momentum? Surely Eq. (17-9) shows that the old form, which assumed that mass is independent of velocity, is no longer strictly tenable.

In the light of the mass-energy equivalence, we can now examine several other nuclear processes.

(2) *Rutherford's nuclear transmutation.* In 1919 the great British physicist Ernest Rutherford observed a then very strange occurrence. During the bombardment of nitrogen gas with alpha-particles from a radium preparation, there were produced during the collision some high-speed particles, far longer in range than the stream of incoming alpha-particles. After detailed investigation, Rutherford proposed that the following was happening: The alpha-particles, known to come in with a kinetic energy  $KE_\alpha = 12.3 \times 10^{-6}$  erg, collided (inelastically) with the nuclei of relatively stationary nitrogen atoms. At that instant there would occur a *transmutation* of the nitrogen nucleus into a rare but well-known type (or isotope) of oxygen, observed to recoil with a kinetic energy  $KE_o \doteq 1 \times 10^{-6}$  erg; this transmutation would be accompanied by the simultaneous ejection from the changing nucleus of a positively charged particle, a *proton* with  $KE_p \doteq 9.5 \times 10^{-6}$  erg. Recognizing that alpha-particles and protons are respectively the same structures as the nuclei of ordinary helium (He) and hydrogen (H), we can write this nuclear reaction in the usual symbolism (where the superscripts 14, 4, 17, and 1 refer to the "mass numbers," i.e., the approximate masses in amu, and where the more accurate masses, as later separately and carefully determined for those particles *at rest*, are entered as a last line. These rest masses refer to the inertia which the particles will exhibit if they do not also have kinetic energy.)\*



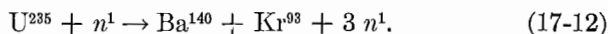
Rutherford's suggestion was epochal: it was the fulfillment of the ancient alchemic dream in unexpected form, the first induced transmutation of matter, although, of course, in quantities too small to be directly recovered, and it initiated a whole new way of thinking in nuclear physics. Obviously the total rest mass of  $\text{O}^{17}$  and  $\text{H}^1$  nuclei is larger than the rest mass of  $\text{N}^{14}$  and  $\text{He}^4$ , precisely by 0.0012 amu; and on the other hand, the kinetic energy of the incident particle is larger (by about  $1.8 \times 10^{-6}$  erg) than the sum of kinetic energies for the recoiling particles  $\text{O}^{17}$  and  $\text{H}^1$ . We expect that the

\* Recall that 1 atomic mass unit =  $1.6602 \times 10^{-24}$  gm. The numerical values given here are recent and were not available to Rutherford with this accuracy. As is customary, the calculation is made on the basis of the mass of the *neutral atom*, i.e., the nucleus plus the electrons around the nucleus. Why is this permissible?

term  $\Sigma(m_0 + KE/c^2)$  will be constant before and after transmutation, and this is true. (Check this.) The observed decrease in kinetic energy is compensated by the increase of rest mass.

(3) *Hydrogen fusion.* The previous historic reaction, in analogy to some chemical processes, may be called *endothermic*. Other nuclear processes can be *exothermic*, energy-developing. When the nuclei of hydrogen are so agitated (as by the heat in stars or in "atomic" explosions) that they can collide with one another despite their mutual electric repulsion, then they may presumably join or fuse, and thereby two hydrogen nuclei (of the "heavy" variety) may become one nucleus of helium; the latter has, however, less mass than that of both building blocks together, and the difference is evidenced as an increase in the kinetic energy available to the end products of the fusion process.

(4) *Uranium fission.* In Section 16-6 we presented one reaction scheme for the fission of the isotope of uranium  $U^{235}$ :



This, too, is an exothermic process, i.e., on the right side of the equation, the sum of the rest masses is less and the sum of the kinetic energies greater than on the left. At the same time, the emission of several neutrons in fission makes it possible for one such nuclear process to trigger off several like it in the immediate neighborhood, and as each exothermic reaction generates more of the same, an explosive chain reaction is created.

PROBLEM 17-31. Following are the data for Eq. (17-12). The rest masses of  $U^{235}$  and the neutron are 235.116 and 1.0090 amu respectively. The energy of the incident neutron is negligible. After fission, the fragments have a total of about  $3.2 \times 10^{-4}$  erg of kinetic energy. (a) What is the total rest mass of the fragments? (b) What is the total inertia of the fragments as it would be measured right after fission, before they have handed on their kinetic energy by collision with other atoms?

PROBLEM 17-32. As first explained by H. A. Bethe (1939), the sun obtains its energy by a fusion process that turns four hydrogen nuclei into one helium nucleus in a cycle involving the catalytic action of other nuclei (oxygen, nitrogen, and carbon) and the emission of energy in different radiations. If each of the hydrogen atoms has a rest mass of 1.0081 and each helium atom a rest mass of 4.0039, calculate the energy available in each fusion cycle. (Note: The process is so slow that only 1% of the sun's hydrogen is consumed in  $10^9$  years; still, many millions of tons of mass are lost each minute by radiation of energy into space.)

(5) *The neutrino problem.* When the nucleus of a radioactive element "decays," it does so by emitting a particle, perhaps a  $\beta$ -ray (an electron) or an  $\alpha$ -particle (a helium nucleus), often together with a  $\gamma$ -ray (high frequency x-ray).<sup>\*</sup> For example, when an atom of  $U^{235}$  is *not* bombarded but allowed to decay peacefully, it eventually emits a fairly high-speed  $\alpha$ -ray with very specific kinetic energy. And counting in the energy of the accompanying  $\gamma$ -ray, there is a perfect balance between the mass of the uranium nucleus before disintegration and the sum of the rest masses and the mass-equivalent of the energies after disintegration, just as we now would expect.

But things are less simple if we analyze what happens during those radioactive decays in which an electron is emitted ( $\beta$ -decay), e.g., in a species of Bismuth,  $Bi^{214}$ . The  $\beta$ -ray is clearly observed to leave with an average energy far below that necessary to account for the resulting over-all deficit of rest mass; specifically, the deficit corresponds to a mass 1.7 times that of the rest mass of an electron, or the energy-equivalent of about  $1.4 \times 10^{-6}$  erg. At once some explanations suggest themselves. Perhaps a  $\gamma$ -ray of that energy is emitted also; but we cannot find it on testing the hypothesis. Perhaps another particle is also emitted, either one with a mass about that of the electron and little kinetic energy, or one with a much smaller mass and correspondingly larger energy. But the cloud chamber and similar equipment show no tracks such as would be left by the hypothetical particle if it were charged—quite apart from other reasons why we do not expect the emission of another charged particle. Very well, perhaps the hypothetical particle is *uncharged*, neutral like the neutron, but of course necessarily much smaller, for the neutron has a mass of about 1840 electrons. (This hypothesis stems from the work of Wolfgang Pauli (1933) and of Enrico Fermi (1934) who named the elusive particle *neutrino*, the Italian diminutive for neutron.)

Further experiments lead to a strange impasse. Even a small, neutral high-speed particle ought to be detectable in some way; yet the neutrino seemed to show no other effects than the ones for which it was introduced—to account for the missing quantity. This is hardly a firm basis for reality, and it would be at least equally logical to announce that in  $\beta$ -decay the law of conservation of mass and energy breaks down. Surely science has learned long ago that gen-

---

<sup>\*</sup> Note that it is customary to call anything emitted from the nucleus "radiation," whether it is actually in the form of material particles ( $\alpha$  and  $\beta$ ) or radiant energy ( $\gamma$ ).

eral laws do eventually reveal limits, that it is folly to cling to a conceptual scheme if in order to "save the phenomena" it must be patched up by independently unconfirmable hypotheses.

But in truth, the case here was more complex, both worse and better than has so far appeared. For not only are mass and energy not conserved in  $\beta$ -decay, neither is momentum! More specifically, nuclei have spinning motion, like the sun in our solar system, and with it angular momentum. We expect that the total angular momentum of the decay products (residual nucleus plus  $\beta$ -particle) is equal to the angular momentum of the original or "parent" nucleus. Experiments say clearly: Not so. We add to the list of missing quantities a measurable deficit of angular momentum; if the neutrino exists, it must spin. But, in fact, so do all known fundamental particles; and so we might say that the new disaster really confirms a little the plausibility of the hypothesis. Moreover, the linear momentum of the recoiling particles after decay also lacks a certain amount from the balance predicted by the LCMm. Thus we have the choice of giving up all conservation laws at this point, or else of letting *one* mechanism, the neutrino, save them all in one stroke.

The temptation to follow the latter alternative is increased by additional evidences that the neutrino hypothesis is useful beyond its original intent. In some radioactive decay processes, a positive electron (positron) is emitted. The same deficiencies are observed as in  $\beta$ -decay, and the same explanation will serve. And in time, other nuclear processes have become known in which the neutrino hypothesis would save the conservation laws. Although every experiment seems to make it more plain that the rest mass of the neutrino must be extremely small, if not indeed zero, yet this may be but another instance of a phenomenon so well illustrated by the history of physical science, where the difficulty in great part is one of conceptualization, where, in short, our previous intuitions based on the mechanics of large objects must be disciplined by long familiarity with a new type of fact.

But the answer still lies in the future. Our reason and our scientific faith tell us that if the neutrino hypothesis is not proved right, the alternative will have to be at least as consistent and useful; surely, the alternative cannot be the *lack* of a physical law.

**17-8 The mechanical equivalent of heat. Mayer and Joule.** There remains a final extension of the concept of energy: to problems involving *heat*. Historically it was just this extension, achieved in

the middle of the last century, which in fact clarified and brought together the several separate energy concepts, and so made possible the first general law of conservation of energy.

The reason for delaying this part of the story so long is that the details are thoroughly mixed with a number of different developments. Even at this stage it will be best to reserve for the next chapter the main outlines of the theory of the nature of heat itself, and to concentrate here on the aspects immediately associated with the LCE. So let us tentatively define heat as the energy that can change the temperature or the physical state (solid, liquid, gaseous) of a material body. How heat energy can do such things is not to be considered just now; we shall start with the empirical fact that it can do so. Next, how shall we measure heat? Here we must adopt an old convention: that quantity of heat is arbitrarily called *unity* which raises the temperature of a one-gram mass of water through just one degree Celsius ( $1^{\circ}\text{C}$ ). The name for that unit of heat, whether supplied by means of a flame, an electric heater, or otherwise, is the calorie (cal), so one may say that it takes 1600 cal to heat 20 gm of water from room temperature ( $20^{\circ}\text{C}$ ) to the boiling point ( $100^{\circ}\text{C}$ ).<sup>\*</sup> Conversely, if we should be required to supply a certain body, say a block of steel, with exactly one calorie of heat, we should take 1 gm of hot water, bring it in contact with the block, and wait until the temperature of the water has dropped  $1^{\circ}\text{C}$  by conduction to the metal (making sure that there are no other heat losses).<sup>\*\*</sup> Or if we wish to rid an object of a measurable amount of heat, we might place some ice upon it to see how much is melted, having independently ascertained that melting ice takes 80 calories per gram of ice (a quantity referred to as the *latent heat of melting*). This method found much favor in early experiments.

After one experimentation one may then assign to every other substance a *specific heat* ( $s$ ), a value that is numerically equal to the heat energy needed to raise the temperature of 1 gm of that substance by  $1^{\circ}\text{C}$ . For water,  $s$  is 1 (cal/gm)/ $^{\circ}\text{C}$ , by definition; for copper, it is about 0.09 (cal/gm)/ $^{\circ}\text{C}$ ; for lead, 0.03 (cal/gm)/ $^{\circ}\text{C}$ ; and for most other solids and liquids it has some intermediate value.

In the case of gases—and here we may include air, a mixture of

---

<sup>\*</sup> Strictly speaking, the heat needed per gram per  $^{\circ}\text{C}$  varies a little with the actual temperature, but that can be neglected here.

<sup>\*\*</sup> *Problem:* suggest in detail a more practical way of supplying known quantities of heat. Suggestion: calibrate a steady source of heat in terms of its effect on water.



gases—there is an interesting and important difficulty. The specific heat depends to a marked degree on the method of measurement. If, for example, we take a sample of air in a tight enclosure at a constant volume, we must supply 0.17 cal per gm per  $^{\circ}\text{C}$  temperature rise and, incidentally, we shall note that the pressure in the gas rises. Now we repeat this experiment with the same sample in a cylinder having one freely movable side or piston (Fig. 17-24). The pressure inside and outside the cylinder will always be the same, for the inflow of heat to the gas can now be attended

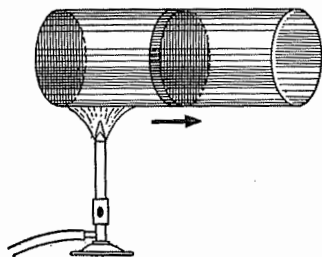


FIG. 17-24. Heating a sample of gas in a cylinder with freely sliding piston.

by an expansion, and it will be found that one needs 0.24 (cal/gm)/ $^{\circ}\text{C}$ , i.e., 40% more than before. We must specify two specific heats for gases, one for constant volume, another for constant pressure.

This reminds us also that our definition of heat was too limited. Heat not only changes temperature and physical state, but it can also move objects (here, a piston) and so can be made to produce work. This was, of course, a commonplace after the development of the steam engine, but the relation between heat and work was by no means clear before the 1850's. Up to that time, it was believed that the flow of heat itself determined the motion of the parts of a heat engine in the same manner as the flow of water over a water-wheel makes the mill turn, and that heat itself was no more consumed in the process than is water in the other case.

Between that point and a truer understanding of the phenomenon there had to develop three major ideas: the idea that heat can be created by, and at the expense of, other forms of energy, e.g., by mechanical work and by electric currents; second, the idea that this exchange can proceed the other way also, i.e., that heat is a form of energy that can be converted into other forms, so that heat is "lost" in exact proportion to work done or electrical energy created; and third, the idea that this general two-way interchangeability of all energies proceeds according to a very strict quantitative equivalence: always  $x$  calories of heat for  $y$  ergs of other energy, and  $y$  ergs for  $x$  calories. Once all this was established by experiment, even for special cases of energy conversion, the general law of conservation of energy could be stated and believed. Before that, only some very

specific versions of the LCE for conservative systems and purely mechanical problems were imaginable and profitable, as for example the conservation of *vis viva* for elastic collision, or the law that  $\Delta KE + \Delta PE = 0$  for the frictionless pendulum.

These crucial ideas grew up roughly in three corresponding stages, although their details are not at all separable. First we may mention the experiments of Count Rumford (1753-1814), that amazing American who began life as Benjamin Thompson of North Woburn, Massachusetts, and by 1798, a refugee from his native country, had been knighted by King George III and was in the employ of the Elector of Bavaria as Chamberlain, Minister of War, and Minister of Police, still finding time from his many duties and schemes to form some capable and historic experiments.\* In one series he made some rather qualitative measurements of the heat produced by the expenditure of mechanical work in friction. Certainly, in itself this was no new discovery—prehistoric man knew how to make fire by rubbing sticks together—but Rumford carried the observations to a point where a simple estimate could be attempted of the numerical equivalence between the work and heat. Specifically, the work against friction was supplied by letting a blunted tool in a horse-driven lathe turn inside a metal cylinder, originally part of a casting for a gun but “designed for the express purpose of generating heat by friction.” Around the cylinder was built a wooden jacket or box which contained water.

In one of his several experiments, the water, after a start from room temperature, was brought to boiling by the friction in the cylinder in  $2\frac{1}{2}$  hours of continued action. A crude calculation could then be made (although Rumford himself was content with demonstrating how large an amount of heat was developed there). The total mass of material heated was known, and so was the change in temperature. This gave the total number of calories produced, roughly  $1.2 \times 10^6$  calories using Rumford's data, but taking notice that “no estimate was made of the heat accumulated in the box [the wood itself], nor of that dispersed during the experiment.” The work done to obtain this heat was judged by the fact that one horse could “easily” turn the machine. On the basis of experiments, it had been suggested earlier by James Watt that for purposes of calculation one horse should be taken to supply  $7.5 \times 10^9$  ergs of work per second (= one “horsepower”). Thus in  $2\frac{1}{2}$  hours, the work done on the cylinder could be  $6.7 \times 10^{13}$  ergs. It might also well have been less,

\* No one should miss the essay on Rumford's life and work in John Tyndall's *New Fragments* (1896)

either if the horse was not worked to capacity or if there were some losses to friction in other parts of the machine. But on taking the ratio of work done to heat evolved—a ratio that has come to be called the *mechanical equivalent of heat*, symbolized by  $J$ —then the sources of error in both factors would tend to counteract each other, and the figure obtained is

$$J = \frac{6.7 \times 10^{13} \text{ ergs}}{1.2 \times 10^6 \text{ cal}} = 5.6 \times 10^7 \text{ ergs/cal.}$$

If not precisely correct, this should at least be of the right order of magnitude. But mark well: it has not been shown that  $J$  is a universal constant, e.g., that there is an exact and fixed rate of exchange between heat and work in many different situations, nor has the heat equivalent of forms of energy other than work against friction been mentioned so far, nor has the converse convertibility from heat to work. In truth, Rumford's main goal in these experiments was along different lines, as we shall see in the next chapter.

But several other contributions at that time pointed also to a relationship between heat and other forms of energy. The conversion of kinetic energy to heat, as by the quick hammering of a nail on an anvil, was, of course, common knowledge; so was the liberation of heat in chemical reactions. About 20 years before Rumford, Lavoisier and Laplace had introduced a first note from physiology by their experiments to prove that a living guinea pig developed animal heat at about the same rate as does the chemical process of burning its food intake. In 1819, the French chemist and physicist, P. L. Dulong, showed that on compressing gas quickly the heat developed in the gas is proportional to the mechanical work done upon it. Research was begun on the heat produced by electric currents.

And all the while, two other influences develop. The energy concept becomes more clearly defined; the word "energy" itself (in its correct meaning) dates from 1807. And as particularly illustrated in the work of the British physicist Michael Faraday on the interaction of electric current, magnetism, and light, there spreads increasingly a belief that the several types of natural activities are all connected and are ultimately all part of one unified story. In reviewing this period prior to the 1840's one is struck by a widespread groping toward the general law of conservation of energy from many different sides, often far beyond the context, the available evidences, or the clarity of the concepts themselves.

Of the three great ideas leading to the general LCE, the second seems to have been stimulated by the practical example of the

mechanism which most prominently showed the close relationship between heat supplied and work done, namely, the steam engine. Particularly in France, there was much early interest in the effects of heat on gases and on the work done by them. In an investigation that later assumed particular significance, the chemist L. J. Gay-Lussac showed that a sample of gas will expand freely into the previously empty part of a set of connected vessels *without* measurably changing its over-all temperature, whereas expansion against some surrounding medium is possible without temperature drop only if heat is supplied.

These and several other ideas about the interchangeability of heat and work were pulled together in 1842 in an essay that for the first time suggested what we have assumed as basic in Sections 17-5 to 17-7, i.e., a general



FIG. 17-25. Julius Robert Mayer (1814-1878).

equivalence and conservation of all forms of energy, the equivalence of heat and work being only a special case. This essay, so full of imaginative, almost metaphysical, but largely unsupported generalizations that its publication was at first refused by one of the great journals of science, was the work of Julius Robert Mayer (1814-1878), then a young doctor in the German town of Heilbronn, his native city (Fig. 17-25). He later said that his thoughts first turned to this subject when, as a ship's physician in the tropics, he discovered venous blood to be of a brighter red color than he had expected from experience in cooler climes; and that subsequently, the voyage evidently being long and uneventful, he began to speculate upon the relation between animal heat and chemical reaction, and from there to the interchangeability of all energies in nature.

In his 1842 paper, "Remarks on the Energies\* of Inorganic Nature," Mayer announces:

"Energies are causes: accordingly, we may in relation to them

---

\* In the German original, the old word *Kraft* (Force) is used throughout in this sense.

make full application of the principle—*causa aequal effectum*. If the cause  $c$  has the effect  $e$ , then  $e = c$ ; . . . In a chain of causes and effects, a term or a part of a term can never, as plainly appears from the nature of an equation, become equal to nothing. This first property of all causes we call their indestructibility . . . If after the production of [effect]  $e$ , [cause]  $c$  still remained in whole or in part, there must be still further effects [ $f$ ,  $g$ , . . .] corresponding to the remaining cause. Accordingly, since  $c$  becomes  $e$ , and  $e$  becomes  $f$ , etc., we must regard these various magnitudes as different forms under which one and the same entity makes its appearance. This capability of assuming various forms is the second essential property of all causes. Taking both properties together, we may say, causes are quantitatively *indestructible* and qualitatively *convertible* entities . . . Energies are therefore indestructible, convertible entities.”

This, then, is the first voicing of conservation and equivalence of all energies. But this argument is, of course, false. “Causes” do not have the properties derived by this private logic, nor can one equate energy with “cause” unless there is prior *experimental* proof of the indestructibility and convertibility of energy.

But further on, Mayer makes some applications of his doctrine to chemistry and to mechanics. He continues:

“In numberless cases we see motion cease without having caused another motion or the lifting of a weight; but an energy once in existence cannot be annihilated, it can only change its form; and the question therefore arises, What other forms is energy, which we have become acquainted with as potential energy and kinetic energy [modern terminology], capable of assuming? Experience alone can lead us to a conclusion.”

There follows the argument along expected lines that if heat is generated by the motion of two surfaces against each other with friction, heat therefore must *be* a form of motion—an idea which finds expression throughout the history of science, to which Rumford had previously contributed extensively, and which was to prove correct within a decade—although not on such grounds (see Chapter 20). From Mayer’s fundamental proposition, he concludes further: “If potential energy and kinetic energy are equivalent to heat, heat must also naturally be equivalent to kinetic energy and potential energy.” Consequently, “we will close our disquisition, the propositions of which have resulted as necessary consequences from the

principle *causa aequat effectum* and which are in accordance with all the phenomena of nature, with a practical deduction . . . How great is the quantity of heat which corresponds to a given quantity of kinetic or potential energy?"

This is indeed the crucial point. Mayer's presentation so far has been almost completely qualitative, and indeed rather reminiscent of scholastic science in many ways. This is partly understandable from the fact that Mayer himself had then no facilities for experimentation and found the physicists of his time quite uncooperative. But here, at the end of his paper, deceptively like an afterthought, is finally promised a derivation of the equivalence between heat and work, and by an approach opposite to that of Rumford. Here is the chance for empirical verification.

Mayer's solution is only sketched out, but one may reconstruct the steps. In the circumstances, he had to fashion his illustrative computation from some known data whose significance had by and large not been noted.\* And at that point he selected, as was said 30 years later, "with marvelous sagacity, from existing physical data, the single result on which could be founded a calculation of the mechanical equivalent of heat." Mayer refers to the evidence of *two* specific heats for gases; we recall that 1 gm of air requires 0.17 cal to gain  $1^{\circ}\text{C}$  at constant volume, but 0.24 cal if the piston is free to move and to expand the chamber. Mayer assumes that in the second case also 0.17 cal is needed to increase the temperature, that the rest, 0.07 cal, is needed to move the piston out against the surrounding air (at the constant pressure of 1 atm or  $1.013 \times 10^6$  dynes per square centimeter of piston area). In retrospect, one may see that this idea is inherent in Gay-Lussac's observation that no difference of specific heats exists if the gas can expand into a vacuum, against no resistance. Here, the work done by the gas on the piston (Fig. 17-26) is the perpendicular force ( $F$ ) times the distance it moves ( $s$ ). But since the pressure ( $P$ ) in the gas is defined as force ( $F$ )  $\div$  area ( $A$ ) on which  $F$  acts, therefore the work done on expansion becomes  $(PA) \times s$ . Furthermore,  $A \times s =$  change in volume during the expansion of this 1 gm of air, and that was a known quantity, namely,  $2.83 \text{ cm}^3$ . Thus 0.07 cal of heat appears as 1.013

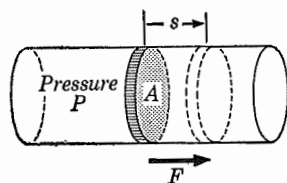


FIG. 17-26.

\* A claim may, however, be made for the work of Marc Séguin (1839), which Mayer then probably did not know of.

$\times 10^6$  dynes/cm<sup>2</sup>  $\times 2.83$  cm<sup>3</sup>  $\doteq 2.86 \times 10^6$  ergs of work, or 1 cal is equivalent to about  $4 \times 10^7$  ergs. (Approximately modern values have been used here; the measurements known to Mayer gave him 1 cal  $\rightarrow 3.6 \times 10^7$  ergs. We recall that, from Rumford's experiments,  $5.6 \times 10^7$  ergs  $\rightarrow$  1 cal. Considering the known uncertainties in the data used by these men, the two equivalence values are quite close.)

For about 20 years, there was virtually no reaction to this work. Scientists quite rightly are wary of such largely qualitative efforts, and this was not a case of a generally recognized conceptual crisis which Mayer's new ideas would have resolved. But there is something to be said on the other side; scientists are by no means always immediately perceptive of the real, long-range value of an unorthodox contribution. For Mayer did develop his principle of general indestructibility and strict convertibility of energy further in several subsequent essays, sometimes presenting good references to proof by calculation, and extending the principle "with remarkable boldness, sagacity, and completeness" to chemistry, astronomy, and living processes. But although the work in retrospect was clearly one of genius and "on the whole, a production of extraordinary merit," Mayer continued to be so utterly rejected that his mind, overtaxed by the strain of his work and his lonely fight, gave way for a time. He recovered and witnessed growing recognition and honor in the closing years of his life.

This change of personal fortunes was by no means a sign that science had become sentimental about Mayer's condition; rather, it was the consequence of a general acceptance of the LCE on the basis of an entirely different series of events, above all, the remarkable and persistent experimental work of a British amateur of science, James Prescott Joule (1818-1889). The distinction between the two men has never been put better than by John Tyndall\* in his still very readable classic *Heat Considered as a Mode of Motion* (1863 and later editions):

"True to the speculative instinct of his country, Mayer drew large and weighty conclusions from slender premises, while the Englishman aimed, above all things, at the firm establishment of facts. And he did establish them."

Joule was born near Manchester, the son of a well-to-do brewer to whose business he succeeded. But his dedication from the start

\* From whom stem also the quotations in the previous two paragraphs.

was to science. At 17, he was a pupil of the chemist John Dalton, and at 22, he had begun a series of investigations that was to occupy the greater part of his life: the proof that when mechanical work is turned to heat under any circumstance, the ratio of work done to heat evolved has a constant and measurable value. In part, his motivation seems to have been essentially the same as that of every other proponent of a conservation law; in his words, to him it was "manifestly absurd to suppose that the powers with which God has endowed matter can be destroyed."

The most powerful drive in Joule was undoubtedly the sheer love for experimental investigations.\* Independent of Mayer's work (with which he said he was "only imperfectly acquainted" as late as 1850), he announced in 1843 the first result of his labor, the comparison of mechanical work needed to operate an electric generator and the heat produced by the current so generated:  $J = 4.51 \times 10^7$  ergs/cal. This was an average of 13 experiments with fairly large uncertainties, for in all such experiments it is extremely difficult to obtain precise and meaningful data, because of unavoidable leakage of heat and the like. Later the same year, Joule measured the relation between heat of friction for water flowing through thin pipes and the work needed to produce the flow ( $4.14 \times 10^7$  ergs/cal), and he concludes:

"We shall be obliged to admit that Count Rumford was right in attributing the heat evolved by boring cannon to friction . . . I shall lose no time repeating and extending these experiments, being satisfied that the grand agents of nature are, by the Creator's fiat, *indestructible*; and that whenever mechanical force is expended, an exact equivalent of heat is *always* obtained."

But Joule certainly did not rest now. Despite the coolness with which his work was generally received, he continued with almost obstinate zeal. In 1844 he measured the ratio of work needed to compress a gas to the heat so created ( $4.27 \times 10^7$  ergs/cal), in principle rather close to the method which Mayer had considered in 1842. A little later, Joule made the first of a long series of more and more accurate experiments to find the mechanical equivalent of heat by friction, e.g., by stirring water with paddlewheels in a heat-insulated container. This is initially a delicate experiment because the temperature rise is fairly small, and his result ( $4.78 \times 10^7$  ergs/cal)

\* Yet we must also not underestimate the occasional *practical* motivations. Joule initially was very interested in attempts to secure power for his plants more cheaply.



is markedly different from the others. By 1847 the technical difficulties were in hand; stirring water and also sperm oil gave very closely the same result ( $\sim 4.20 \times 10^7$  ergs/cal).

When he wished to report this particular experiment at a scientific meeting in 1847, a crisis arose. Joule was requested by the chairman to confine himself to a "short verbal description," partly because his previous communications, as Joule himself recalled later, "did not excite much general attention." "This I endeavoured to do, and, discussion not being invited, the communication would have passed without comment if a young man had not risen in the Section and by his intelligent observations created a lively interest in the new theory. The young man was William Thomson" (afterwards raised to peerage as Lord Kelvin), then 23 years old and destined to become one of the foremost British scientists, collaborator and staunch supporter of Joule's views.

The year 1849 brings further results on friction: in water, in mercury (where the temperature change is larger by virtue of its lower specific heat), and for two iron disks rubbing against each other. Joule compares his methods and results with those of Rumford, which they follow in principle; it is Joule who makes the sample calculation of  $J$  on page 349. Then he summarizes all his previous work and assigns to the mechanical equivalent of heat the value  $J = 4.15 \times 10^7$  ergs/cal. Today, as a result of a century of increasingly refined precision methods, the value generally used is  $4.18 \times 10^7$  ergs/cal, and  $J$ , of course, stands, with greatest justice, for "Joule's Equivalent."

By 1850, after 10 years of devoted activity, Joule's work began generally to be taken more seriously, the gratifying result of a well-planned experimental assault. During the next 18 years, he carried through three further determinations of  $J$ . Without a doubt, it was the sum of his consistent experimental work that made acceptable the idea of an enlarged, general law of conservation of all forms of energy in place of the previous specific and limited application of the law to conservative mechanical systems alone. But, as usually happens, Joule's triumph depended in large measure on developments beyond his own participation. It will be seen in the next chapters that at about the same time the general thinking on the nature of heat and of matter itself was undergoing profound changes toward a view more favorable to the conservation of heat as a form of energy, namely, the kinetic energy of motion of the molecules. This trend removed the cause of a great deal of the original resistance to Rumford, Mayer, and Joule.

And secondly, the idea of a general LCE found powerful and eloquent champions, particularly at first the young German physiologist and physicist, Hermann von Helmholtz.\* He did what Mayer, then still in oblivion, had not quite done and Joule had never attempted, namely, to show by mathematical demonstration the precise extent of the conservation law to various fields of science—mechanics, heat, electricity, magnetism, physical chemistry, and astronomy—to derive by its aid, in terms that the professional scientist would at once recognize, explanations for old puzzles and new—confirmable mathematical relations. Indeed, as the LCE gradually won more adherents and wider application, it was discovered that Mayer's vision had been prophetic, that the LCE provided striking general connections among the various sciences. It was perhaps the greatest step toward unity in science since Galileo and Newton, and a powerful guide in the exploration of new fields. It has been well said, "Because of its practical use, and for its own intrinsic interest, the principle of the conservation of energy may be regarded as one of the great achievements of the human mind" (*Dampier*).

A footnote to this short historical account may be in order to round out the not unrepresentative sketch of the growth of a conceptual scheme in science with one other rather typical detail. As Mayer's and Joule's work finally won partisans, there developed a sometimes rather bitter fight as to which of the two had been the real originator of the general doctrine of conservation of energy in its various forms. The Joulites disparaged Mayer's contribution of 1842 that gave him technical priority over Joule, whose first but quite independent paper was dated 1843; and Joule himself wrote in 1864:

"Neither in Séguin's writing (of 1839) nor in Mayer's paper of 1842 were there such proofs of the hypothesis advanced as were sufficient to cause it to be admitted into science without further inquiry . . . Mayer appears to have hastened to publish his views for the express purpose of securing priority. He did not wait until he had the opportunity of supporting them by facts. My course, on the contrary, was to publish only such theories as I had established by experiments calculated to commend them to the scientific public, being well convinced of the truth of

---

\* Helmholtz was then 26 years old. It is probably no coincidence that most of the protagonists were quite young, under 30, when they became interested in the LCE.

Sir J. Herschel's remark that '*hasty generalization is the bane of science.*' "

On the other side, and often not much more fairly, those partial to Mayer's claim to priority would say with Helmholtz:

"The progress of natural science depends on this: that new [theoretical] ideas are continually induced from the available facts; and that afterwards the consequence of these ideas, insofar as they refer to new facts, are experimentally checked against actuality. There can be no doubt about the necessity of this second endeavor; and often this second part will cost a great deal of labor and ingenuity, and will be thought of as a high accomplishment of him who carries it through well. But the fame of discovery still belongs to him who found the new idea; the experimental investigation is afterwards a much more mechanical type of achievement. Also, one cannot unconditionally demand that the inventor of the idea be under obligation to carry out the second part of the endeavor. With that demand one would dismiss the greatest part of the work of all mathematical physicists."

Here we see an old fight, or rather two fights: one for priority (and each great idea, after it is accepted, tends to give rise to such a dispute) and second, the fight concerning the relative merits of the experimentalist and the theoretician. In connection with both we get another glimpse of the human element behind the printed equation.

**17-9 General illustrations of the LCE.** As has been demonstrated, the LCE may take on any one of many different forms, depending on the intended application. But broadly speaking, we have met three main groups of formulations:

(i) *For an isolated system*, the sum of all forms of energy remains constant (the net changes add up to zero) although the internal energy may take on different forms. Examples are collisions between balls or molecules, and disintegration in natural radioactivity.

(ii) *For a system to which energy is given*, including work done or heat supplied to it, but which does not in turn give energy to the world outside the system, the energy supplied to the system is equal to the total change in internal energy of all forms. Examples are a block being pulled by a string, gas in a cylinder being compressed by a piston, induced transmutation, and fission by neutron bombardment.

(iii) *For a system to which energy is supplied and which also does work on or loses energy to the outside world*, the energy supplied to a system is equal to the total change of internal energy in the system.

plus the energy given up by the system. Examples are machines and Joule's paddlewheel experiment with heat leakage.

But of course the LCE is a *single* law, and the forms it takes depend only on how large we choose the "system" under consideration to be. Thus when a man hoists a load by means of a pulley, the form (iii) applies if the pulley itself is the whole system, and the form (i) if the earth together with man, pulley, and load is considered one system. Furthermore, the first two (and one other possible form: which? examples?) are but special examples of the third form, with one or the other factor set equal to zero.

In scientific writings one encounters the LCE phrased in many different ways. One equivalent name, "First Law of Thermodynamics," is the LCE in such form as applies in particular to cases where exchanges between heat and mechanical energy enter. Another equivalent formulation is of particular interest: it is the statement that a *perpetuum mobile* (perpetual mover) is impossible; for by *perpetuum mobile* is generally meant a machine that has an efficiency of *more* than 1.00 (100%), one which yields more output than input. One may even identify the rise of the LCE in mechanics with the continued unsuccessful attempts to make such a machine, for it should be remembered that this project was very popular indeed for hundreds of years. The search, which engaged some of the greatest intellects, was in itself no more implausible than the earlier, parallel, and equally fruitless search for the Philosopher's Stone, and it reflected the preoccupation of an increasingly more machine-minded civilization. However, it probably points up equally that most human of all traits, the wish to get something for nothing.

The inability to make a perpetual motion machine became itself a fact of physical reality, one example of a series of so-called *Postulates of Impotency* which have profoundly influenced physical science. To this class belong the impossibility of transmuting matter chemically, which helped to shape the idea of chemical elements; the inability to discover the relative motion of the ether in optical experiments, which was turned into a fundamental postulate of Relativity Theory; the impossibility of obtaining net mechanical energy from a medium by cooling it below the temperature of its surroundings,\* which is the effective content of the Second Law of Thermodynamics; and several others.

---

\* An engine that thus functioned, if it were not impossible for other reasons, would not be in conflict with the LCE. It is referred to as a *perpetuum mobile of the second kind*, and was once as much sought after as the *perpetuum mobile of the first kind* described above.

Additional familiarity with the LCE in its most general form, and additional evidence of its wide-ranging power will be obtained by the following set of varied examples.

*The human body as an engine.* Once it was thought that animals, including man, must eat only in order to grow and to rebuild used-up tissue. The realization that the food intake itself also provides heat and energy for the functioning of the body was attained only slowly. At the danger of simplifying a complex topic far too much, we may compare the body to a chemical engine: the oxygen we breathe burns the sugars, fats, and some of the proteins from our digested food intake, converting their complex molecules mainly into carbon dioxide, water, and other waste products in an exothermic reaction which liberates heat and, in the contraction of muscles, provides mechanical work.

The balance of energy intake in the form of "combustible" food *vs.* work and heat output has been measured with precision, partly in an effort to check whether life processes also obey the conservation laws in the strict forms that prescribe the activities of inanimate nature. Since the 1890's this problem has been clearly settled in the affirmative. In a representative case, the daily food intake would yield about  $3 \times 10^6$  calories if burned directly, and the heat given up together with the work done and the energy in the waste products come to the same figure within a small experimental uncertainty. But the body, considered as an engine, ordinarily has an efficiency of less than 20%; so if one has to do large amounts of physical work in a short time interval, the attendant heat output will be correspondingly larger, and since the rate of dissipation of heat is not very flexible, the body will heat up drastically.

*Photosynthesis.* If the energy of the animal body were derived from a natural fuel like oil, the food supply on this earth would have been exhausted long ago. The present form of animal life depends on the fact that the body's fuels are continually synthesized by green plants and are readily available, one might say, in the form of either vegetables or vegetarians. The green plant is uniquely able to draw carbon dioxide from the surrounding medium and water, both waste products of animal life, and to recombine the aggregates of carbon, hydrogen, and oxygen, plus those of some mineral salts, into more complex "organic" compounds, each having a large amount of built-in chemical "potential" energy that can be released later during decomposition inside the animal body. For good measure, the plant also releases oxygen into the atmosphere while it builds our food.

For each gram of carbon content, these "fuels" can deliver

roughly  $10^4$  calories' worth of work and heat, and on this depends all animal life except some lower organisms that live by fermentation. But is all this not a crass contradiction of the law of conservation of energy? Do the plant and animal worlds together not constitute a *perpetuum mobile*? The answer is No, because the energy in the manufactured organic molecules is originally taken by the plant from sunlight. The reactions depend on the availability of light energy (and on the presence of chlorophyll and other catalysts); technically this is termed photosynthesis. In this sense, our food is largely sunlight.

If not eaten, the energy-rich plant material in the open will soon decay into its original compounds; but if it is submerged or otherwise suitably kept from ordinary oxidation, it may turn into coal or other heating fuels by slow chemical decompositions that favor the retention of carbon-rich materials.

*The origin of the solar system.* One of the famous old puzzles to which the law of conservation of energy in its early days seemed to give an immediate answer was how the sun could continually emit heat and light at such a staggering rate (about  $3.8 \times 10^{26}$  joules or  $9 \times 10^{25}$  calories every second).\* If this were achieved by burning ordinary fuels or some such chemical reaction, it was difficult to see why the material for it had not been exhausted long ago. Mayer, Helmholtz, and others proposed promising alternatives that the sun's radiated energy was derived from gravitational potential energy by a slow shrinking of the sun or by the capture of comets and meteors. But then it was calculated that the sun's diameter would have to shrink one part out of 10,000 every 2000 years to provide radiation in this manner at the present level. That is too fast to account for the long history of life on earth. Nor can enough meteoric matter be found to give the observed effect. The now accepted source of the sun's energy, namely a nuclear fusion process (see page 343), was of course not imaginable until much more recently. Yet, the original proposal is not altogether without merit, although it failed for the initial application. The same general mechanism of conversion of gravitational potential energy to heat is invoked today in theories concerning the creation of the solar system.

The facts to be explained include the following. At least half of all the stars in our galaxy are binaries (double stars) or stars of higher multiples. These usually revolve quickly about one another. Of the single stars, the very hot ones usually have even more enormous

\* How might this figure have been obtained?

rates of spin, but the "cooler" ones, like our sun, always revolve quite slowly. In our own planetary system, where the orbits of all planets are in the same direction and substantially in the same plane, the total angular momentum of the planets about the sun is very many times that of the sun about its own axis. Now the present-day theories, with different accent on detail to be explained, begin with the same general proposition: the solar system started by the gathering up of a thinly spread cloud of dust, such as the interstellar or "cosmic" dust clouds that are observable with our telescopes. It may be that light-pressure from surrounding stars pushed the dust particles into one another's shadow until the density was sufficient to allow the mutual gravitational forces to predominate.

Such a mass would coalesce in about  $10^9$  years. As the cloud shrinks, any initial rotational motion, if ever so slow, will by the law of conservation of angular momentum increase progressively, and as the gravitational forces pull the particles together, work is done and heat can be generated. According to one line of thought, in the final stages of this process the central portion of the cloud gathers up into one bright, spinning central star (or, in other cases, two or more stars) leaving the outlying and still revolving portions of the cloud to condense into the planets and their satellites.

While the details of this scheme explain also the direction and plane of planetary orbits, the spinning of planets and their moons, and the general distribution of angular momentum, there seem to be theoretical objections in details. To avoid these, the alternative theory, also not yet without its adverse critics, interpolates one sequence: it prefers the fairly complete collapse of the entire initial cloud into one very rapidly spinning star. Eventually the star spins so fast that it breaks into two (or more) sections, or it becomes so hot that nuclear reactions may be set off which explode the star rather in the manner of an "atomic bomb" (*super-nova*). Such an explosion might also separate the original star into two (or more) sections. Whatever the mechanism of separation, the double or multiple stars still spin fairly rapidly at a close distance, but dissipate their energy in and give their momentum to the surrounding dust and debris. In time the rotation of the double stars slows down, therefore they coalesce once more by mutual gravitational attraction, now rotating slowly as one single "cooler" sun, while the rest of the surrounding medium gathers up into planets and satellites. Evidently this scheme orders the different types of stars into an evolutionary sequence; and incidentally allows the speculation that each

exploding star may become the originator of a planetary system. On this basis it has been estimated that the known universe contains perhaps as many as 100 billion planetary systems.

From the present scientific point of view it is natural to think that our earth and all it contains were fashioned by all-pervasive laws that are in continuous operation with similar consequences elsewhere. Here, as a footnote to the powerful and intriguing chapter that the law of conservation of energy has written into science, we cannot forego a brief word of comment. Very probably one theory or another will in time prove that the origin and form of the solar system, including the regularities pointed out by Bode's Law, are explainable by straightforward physical processes. If the past is a guide, this will be interpreted by some as another serious blow of science against religion, for it has generally been held as an axiomatic truth that science would never be able to provide such an explanation, that consequently a theistic assumption was not only necessary but that, indeed, the lack of a scientific alternative was demonstrable proof of the existence of the Deity. Thus, speaking of the planets and comets, Newton himself said in the General Scholium, Book III of the *Principia*:

“But though these bodies may, indeed, continue in their orbits by the mere laws of gravity, yet they could by no means have first derived the regular positions of the orbits themselves from those laws. The six primary planets are revolved about the sun [roughly] in circles concentric with the sun, and with motions directed towards the same parts, and almost in the same plane. Ten moons are revolved about the earth, Jupiter, and Saturn, in circles concentric with them, with the same direction of motion, and nearly in the planes of the orbits of those planets; but it is not to be conceived that mere mechanical causes could give birth to so many regular motions, since the comets range over all parts of the heavens in very eccentric orbits . . . This most beautiful system of the sun, planets, and comets, could only proceed from the counsel and dominion of an intelligent and powerful Being . . .”

The same thinking existed in connection with the problem of the descent of man before it was explained by evolution, with the existence of chemical elements before nuclear physics recently began to supply us with promising theories of their genesis, and with many other examples of this kind. But surely such events teach only this



lesson: it is not only folly but blasphemy to let one's belief in the Deity be based on a premature estimate of what science can *not* do. The opposite course, namely the deification of the discoveries of science, is in truth quite as precarious, for science changes and may come to regard its old knowledge as less relevant. The secure foundation for belief, as all great religious founders have taught, is not the capability of man's mind, or the limitations of his imagination, neither the impressive power of the laws of science, nor the supposed inadequacy of these laws, but faith.

### Additional Problems

**PROBLEM 17-33.** Before Joule obtained a more accurate value of the mechanical equivalent of heat ( $J$ ) he wrote (1845): "Any of your readers who are so fortunate as to reside amid the romantic scenery of Wales or Scotland could, I doubt not, confirm my experiments by trying the temperature of the water at the top and bottom of a cascade. If my views be correct, a fall of 817 feet will of course generate one degree [ $1^{\circ}\text{F}$ ] . . . and the temperature of the river Niagara will be raised about one-fifth of a degree [ $^{\circ}\text{F}$ ] by its fall of 160 feet." (This general prediction he later checked himself, during his honeymoon trip to a large Swiss waterfall.) Now compute from this prediction what value of  $J$  he was using at the time. To convert temperature scales, recall that  $0^{\circ}\text{C} = 32^{\circ}\text{F}$ , and  $100^{\circ}\text{C} = 212^{\circ}\text{F}$ .

**PROBLEM 17-34.** When a 1-kgm ball of lead (with specific heat =  $0.03$  (cal/gm)/ $^{\circ}\text{C}$ ) falls freely through 10 m and is stopped suddenly by collision with the floor, what is the greatest temperature rise you would expect? Cite several causes which will contribute to make the observed temperature rise smaller.

**PROBLEM 17-35.** (a) Calculate the total energy of the earth at one point of its orbit (PE with respect to the sun, plus  $\text{KE}_{\text{rot}}$  in its orbit and  $\text{KE}_{\text{rot}}$  about its axis). (b) Knowing the size and ellipticity of the earth's orbit ( $e = 0.017$ ), find the difference in our actual distance from the sun at perihelion and aphelion, and compute from that plus part (a) the maximum difference in orbital speed.

**PROBLEM 17-36.** In his widely discussed book *Worlds in Collision* (Macmillan, 1950), Immanuel Velikovsky maintains that about 1500 B.C. the planet Jupiter ejected material which, in the form of a comet, passed close to the earth; that the subsequent shower of meteorites falling on the earth stopped its rotation, at least for a short time; that the comet itself eventually turned into the planet Venus; and that later, in 747 B.C. and again in 687 B.C., "Mars caused a repetition of the earlier catastrophes on a smaller scale." Given the mass and radius of the earth, as well as its angular speed of rotation ( $m_e = 5.98 \times 10^{27}$  gm,  $r = 6370$  km,  $\omega = 2\pi$  rad/day), find the angular momentum of the earth which must be carried off before it will stop

rotating and which must be supplied again to start up the rotation afterwards. In the most favorable case the meteorites would have fallen on the equator at grazing incidence, traveling with a speed of perhaps 50,000 mi/hr relative to the surface of the earth. What must have been the approximate total mass of the meteorites? What is the total kinetic energy of rotation which was annihilated by the stopping of the earth and meteorites? What happened to that energy, and how could the earth's rotation start again without another collision? How could gravitational, therefore central, forces between passing spherical bodies radically and suddenly change their state of rotation? What would happen to the oceans and the atmosphere if the earth were to stop rotating by Velikovsky's mechanism?

PROBLEM 17-37. Explain by means of the LCE the following observations: (a) A sample of naturally radioactive material maintains its temperature above that of its surroundings. (b) In a mass of air that is rising to greater height, the content of moisture may freeze. (c) While a battery generates electric energy, it undergoes chemical decomposition at the plates or electrodes.

PROBLEM 17-38. Reread the quotations from Joule and Helmholtz at the end of Section 17-8, and discuss on the basis of the available evidence the fairness of each.

### Further Reading

T. W. Chalmers, *Historic Researches*. London: Morgan Brothers (1949). Chapter II, a very readable case study of the mechanical equivalent of heat.

N. D. Cheronis *et al.*, *The Study of the Physical World*. Boston: Houghton Mifflin Co. (1950). Chapter 34 on Sources of Energy, Chapter 47 on Theories of Cosmogony.

J. G. Crowther, *Men of Science*. New York: W. W. Norton (1936). Study of Joule, also of Faraday, Kelvin, Maxwell, and Davy.

A. Einstein, *Relativity*. London: Methuen and Co. (1920). Also other editions. See also J. W. Knedler, *op. cit.*, pp. 609-637.

B. Jaffe, *Men of Science in America*. New York: Simon and Schuster (1944). Biographies, including Rumford's.

P. Lenard, *op. cit.* (Ch. 9). Biographies of Mayer, Joule, Helmholtz.

E. Mach, *History and Root of the Principle of the Conservation of Energy*. Chicago: Open Court Publishing Co. (1911, original ed. 1872).

W. M. Magie, *op. cit.* (Ch. 2), pp. 196-218. Excerpts from Mayer, Joule, and Helmholtz. Also W. S. Knickerbocker, *op. cit.*, Chapter XXXI.

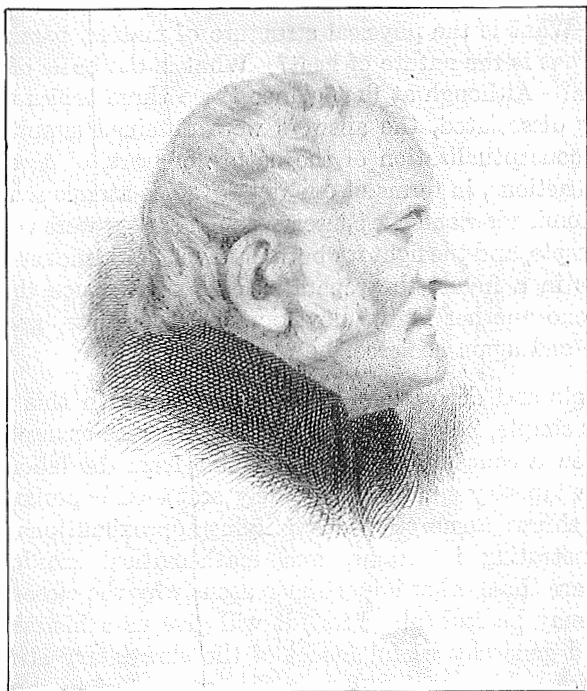
J. Tyndall, *Heat, A Mode of Motion*. New York: D. Appleton and Co. (1893, 6th edition). A classic on the law of conservation of energy and on the kinetic theory. Chapter XVIII reviews Mayer's work.

A. Wood, *Joule and the Study of Energy*. London: G. Bell (1925).

*Part F*

ORIGINS OF THE ATOMIC THEORY  
IN PHYSICS AND CHEMISTRY

- 18 The Physics of Gases and of Heat
- 19 The Atomic Theory of Chemistry
- 20 The Kinetic-Molecular Theory of Gases



*John Dalton (1766–1844)*

## ORIGINS OF THE ATOMIC THEORY IN PHYSICS AND CHEMISTRY

The story of the development and the gradual acceptance into science of the atomic view of matter is astonishing in many ways—in its origins, in the length and vigor of the debate, in the diversity of men and fields involved, in the often quite unexpected concurrence of separate arguments, and in the ever-growing flood of consequences that has finally swept over all fields of study.

The account in this section will carry the development to about the end of the 19th Century. In retrospect it is clear that the model of the atom at that point had grown out of three separate types of questions: What is the physical structure of matter, particularly of gases? What is the nature of heat? What is the basis of chemical phenomena? Although at first glance these three topics may seem completely dissociated, the answers were obtained simultaneously, in a *joint* conceptualization of the nature of heat, of gases, and of chemical reactions, in terms of one quantitative atomic scheme. In fact, the atomic view and the conservation of energy were very closely linked concepts, and the fight for both involved the same set of protagonists. As in a fugue, the separate themes introduce themselves, approach one another in the development, superpose, grow apart, and in the end again coalesce.

The concepts and derivations we shall encounter in this story are not always simple, but there exists hardly a better example on this level to show a conceptual scheme evolving from the labors of generations—a tapestry spun out of widely scattered experimental observations, shrewd assumptions, simplifying approximations, detected errors, penetrating intuitions, and mathematical manipulations. And there are three other important reasons why the closer study of this topic may be fruitful. First, it will give us a historically important and generally useful model of the elementary structure of matter. From this model, and from the outburst of successful speculations which it generated, there arose much of the physics and chemistry of the period immediately preceding the 20th Century—and its failures were the stimuli for much in the new physical science of our own time. Second, not only is the kinetic theory a connecting link between the Newtonian and the contemporary approaches to physics, but it leads from *macrocosmic* physics to the world of *sub-microscopic* phenomena. And finally, this topic will introduce us, though of necessity rather sketchily, to the important type of physical law which, instead of *certainly* and the prediction of individual events, deals with *probability* and the discussion of a multitude of events.

## THE PHYSICS OF GASES AND OF HEAT

**18-1 The nature of gases—early concepts.** From the time of the earliest recorded philosophical speculations to this day, the mind of man has been haunted by some paradoxical propositions: On the one side we can with our hands cut or subdivide gross matter—a rock or a quantity of water—into smaller and smaller parts, until the crudeness of our tools or the deficiency of our sight, *but never the material itself*, calls a halt to our experiments. Consequently, we might conclude that matter is inherently infinitely subdivisible. But on the other hand, our imagination refuses to function properly when we demand that it present to us matter as a collection of truly infinitely small parts, and this view makes for logical problems as well. So where our senses tend to one view of matter, our reason persuades us of another.

The great compromise was formulated about 2400 years ago by the Greek philosophers Leucippus and Democritus, and later extended by Epicurus. It is of course the atomistic idea. Our senses and our reason would both rest satisfied if we assumed that matter is indeed divisible far beyond immediate experience, but that there does exist an ultimate substructure of infinitely hard, uncuttable, indivisible particles, which we might provisionally name *corpuscles* or *atoms*.\*

The Roman poet Lucretius, whose *De Rerum Natura* was cited before in connection with the origins of the law of conservation of mass, exemplified the teaching of those now largely lost Greek original works, and gave a summary of this atomistic science that “marks the climax of Roman scientific thought” (Sarton)—although we have noticed, of course, that this work is not to be regarded as primarily a science text.

His leading theme is “All nature then, as it exists by itself, is founded on two things: there are bodies and there is void in which these bodies are placed and through which they move about . . .” Ordinary matter is composed of these two realities: solid, everlasting particles, and, on the other hand, the void, or what we might loosely call the vacuum. You will notice that by their very definition these atoms are not directly perceptible, and until about 150 years ago there were very few fruitful and unambiguous consequences that

---

\* From the Greek *atomos*, indivisible.

could be drawn from the atomic hypothesis to serve as experimental confirmations of this challenging speculative idea. Even at the turn of this century, when such confirmation was at hand, there were important scientists who still rejected the concept of atoms as too convenient a fable, unworthy of serious consideration by "hard-boiled" investigators.

It is therefore not surprising that the structure of matter was for a long time the subject of inconclusive discussions, whether the atomistic view was rejected (as by Plato, Aristotle, and their diverse followers) or in some form accepted (by Galileo, Gassendi, Bacon, Descartes, Newton, Leibniz, and many of their contemporaries). We might speculate that the idea of small particles became more convincing on the basis of Newton's work where, as we have seen, the postulates leading to the law of universal gravitation provided that the *particles* of matter are the agencies of mutual attraction. Of course, these are not necessarily identical with ultimately *indivisible* particles or atoms. Yet the surpassing success of the treatment of gravitation in terms of the mechanics of particles could turn the thoughts again to the idea that the mechanics of small discrete bodies is the key to all phenomena. Newton himself wrote in the introduction to the *Principia*:

"Then from these [gravitational] forces, by other propositions which are also mathematical, I deduce the motions of the planets, the comets, the moon, and the sea. I wish we could derive the rest from the phenomena of nature by the same kind of reasoning from mechanical principles; for I am induced by many reasons to suspect that they all may depend upon certain forces by which the particles of bodies . . . are either mutually impelled to one another and cohere in regular figures, or are repelled and recede from one another; which forces being unknown, philosophers have hitherto attempted the search of nature in vain, but I hope the principles here laid down will have brought some light either to that or to a truer method of philosophy."

Later, in the *Opticks* (published 1704), we find this visionary, yet still unformalized opinion on atomism:

"All these things being considered, it seems probable to me that God in the Beginning formed Matter in solid, massy, hard, impenetrable, movable Particles, of such Sizes and Figures, and with such other Properties, and in such Proportion to space, as most conduced to the end for which he formed them; and as these

primitive Particles being Solids, are incomparably harder than any porous Bodies compounded of them; even so very hard as never to wear or break in pieces; no ordinary Power being able to divide what God himself made one in the first Creation. . . . And therefore that Nature may be lasting, the Changes of corporeal Things are to be placed only in the various Separations and new Associations, and Motions of these permanent Particles; compound Bodies being apt to break, not in the midst of solid Particles, but where those Particles are laid together . . .

“God is able to create Particles of Matter of several Sizes and Figures and in general Proportion to the space they occupy, and perhaps of different Densities and Forces . . . Now, by the help of these Principles, all material Things seem to have been composed of the hard and solid Particles above mentioned—variously associated in the first Creation by the Counsel of an intelligent Agent. For it became him who created them to set them in order. And if he did so, it’s unphilosophical to seek for any other Origin of the World, or to pretend that it might arise out of a Chaos by the mere Laws of Nature; though being once form’d, it may continue by those Laws for many Ages.”

Part of Newton’s preoccupation here is theological, and for a most excellent historic reason that we must not pass over if we wish to understand fully the development of the atomistic view. Ever since its early inception, atomism has generally been regarded as atheistic; as the excerpt from Lucretius (Section 15-1) indicated, the atomist openly professed to give an explanation of matter and events *not* in terms of impenetrable designs of an ever-present Creator and Ruler, but by means of the interplay and structure among material bodies. Newton’s solution to the problem of accepting atomism while rejecting atheism, which in extension became the characteristic solution in the mechanistic philosophy of the eighteenth century, was to save a place at the Beginning of Time for the Deity, the designer and creator of the World Machine which from then on by His pleasure could run without continual intervention, and in which all observable events are ultimately explainable by the configurations, sizes, and shapes of the eternal particles or by the mathematical laws of force and motion ruling among them.

In the meantime the atomistic view of matter was extended from an entirely different direction—the research on gases. In 1643 Evangelista Torricelli, once briefly the pupil of Galileo, realized that the atmosphere exerted a pressure on the earth, that in fact we are

living at the bottom of a deep "sea of air." Immeasurably more significant for the life of science was a less sweeping realization—that Torricelli's barometric tube contained a small evacuated space (Fig. 18-1). Galileo too had tried, though in vain, to produce a vacuum, but now, following Torricelli's announcements, this long debated ideal became more practical through the development of pumps by Otto von Guericke in Germany *circa* 1650 and by Robert Boyle in England *circa* 1660 (see Fig. 18-2). Even though

these pumps were relatively crude,\* they did provide a tool for experimenting on gases which had previously appeared elusive, mysterious, and even nonmaterial. With a pump and a glass jar, one could prepare an *experimental space* in which to discern the effects of light, sound, and magnetic or electric forces within an atmosphere rarefied in varying degrees. One could attempt to discover, by the difference of weight of a hollow vessel before and after "exsuction," the weight of the gas within. Conversely, one could also pump gases into a vessel at higher pressures.

**18-2 Digression: the general gas law.** The compressibility of air, realized generally for some time, was put on a quantitative basis through Boyle's celebrated experiment, from which it appeared that for a given mass of gas trapped in a vessel at a constant temperature, any decrease of volume raises the gas pressure proportionally, and vice versa. In modern terminology, we write

$$P \cdot V = \text{constant (at constant temperature),} \quad (18-1)$$

where  $P$  refers to the actual pressure of the gas considered (e.g., in dynes/cm<sup>2</sup>) and  $V$  to the volume of the gas (e.g., in cm<sup>3</sup>). Following Boyle's work, continued effort was made to discover the effect of changes of temperature on the pressure or the volume of a gas.

\* The complete removal of every trace of gas from a closed vessel is impossible today also.

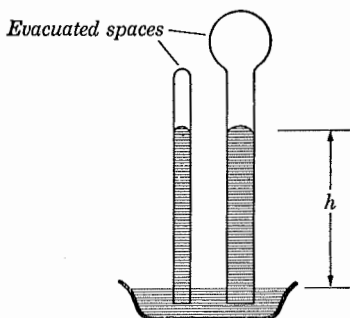


FIG. 18-1. Torricelli barometer (two forms). The mercury column in each vertical glass tube is balanced by the pressure of the atmosphere upon the free mercury surface in the trough. Usually,  $h$  is about 761 mm.





FIG. 18-2. Robert Boyle (1627-1691). Note in the background one of his pumps, set up to evacuate a glass globe.

Toward the end of the 18th century Jacques Charles and, independently, Joseph Louis Gay-Lussac, discovered that if, on heating, it is desired that a constant pressure be maintained in a sample of gas, one must let its volume increase proportionately to the rise in temperature ( $\Delta V \propto \Delta T$  at constant  $P$ ). To visualize this as an experiment, think of a quantity of gas contained in a very thin, easily stretched balloon. The pressure inside and outside is always the same, and constant at one atmosphere. As we now proceed to heat the surroundings ( $\Delta T$ ) the balloon expands ( $\Delta V$ ). What is so very startling is the fact that the fractional increase in volume per degree rise in temperature is exactly the same for all gases, even though their chemical constitutions are different. To appreciate the unexpectedness of this finding, recall that on heating a solid or a liquid one will ordinarily also obtain a length or volume expansion which is proportional to the temperature change, but there is a great difference in the relative amount of volume expansion per degree of temperature between different solids or liquids (about 0.01% for ice, 0.0025% for quartz, 0.02% for mercury and for water, and 0.15% for acetone).

But under normal conditions, almost all gases expand roughly

1/273 (0.37%) of their volume when heated from 0°C to 1°C. This behavior being fairly linear over the ordinary range of temperature, we must conclude that on cooling a gas to 273° below the ice point (i.e., -273°C), its volume should be reduced to zero.\* This value of -273°C is an *absolute zero of temperature* in the sense that generally we do not expect substances to exist there in gaseous form, and that lower temperatures are meaningless as far as gases are concerned. Consequently, we can simplify the law  $\Delta V \propto \Delta T$  (at constant  $P$ ) by writing  $V \propto T$  at constant  $P$  if temperature  $T$  is measured on a new scale of temperature, i.e., the absolute scale (in degrees absolute, °K), where 0°K = -273°C.\*\*

Evidently the rate of exchange between these two scales is

$$\text{temperature (in } ^\circ\text{K)} = \text{temperature (in } ^\circ\text{C)} + 273,$$

a simple enough relationship. For example, 0°C is equivalent to 273°K; room temperature (20°C) is 293°K; water boils at 373°K; and a *one degree interval* on the absolute scale is just as large as a one degree interval on the centigrade scale.

PROBLEM 18-1. Figure 18-3 is a hypothetical plot of experimental data for a sample of gas at constant pressure; here the volume  $V$  is plotted against the temperature  $T$ . The graph illustrates  $\Delta V/\Delta T = \text{constant}$  (Charles' Law). Show by geometrical reasoning that if Charles' law holds, then also  $V \propto T$  (at constant  $P$ ) provided  $T$  is measured on the absolute scale. Show that the proportionality  $V \propto T$  is *not* true if  $T$  is measured in °C.

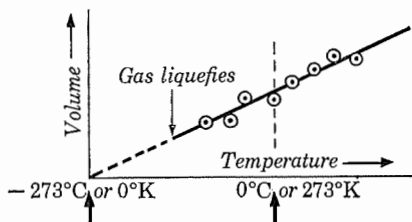


FIG. 18-3.

What interests us most, however, is the obvious next step. If for all gases, by Boyle's law,  $V \propto 1/P$ , and by Charles',  $V \propto T$ , then  $V \propto T/P$ , or

$$\frac{PV}{T} = \text{a constant (usually called } r), \text{ whose value depends only on the sample.} \quad (18-2)$$

\* Or at any rate shrink to an insignificantly small value; actually all gases become liquid before reaching this low limit of temperature and then, of course, cease to obey the law for ideal gases.

\*\* More accurately, -273.16°C. The letter K stands for Kelvin, in honor of Lord Kelvin, prominent contributor to thermodynamics.

We may call the last equation, which is of great usefulness, the *general gas law*, since it relates simply the three variables of interest in any given sample of gas. *Note well the beauty of this law*: once we measure simultaneously three quantities ( $P, V, T$ ) for a gas sample, its whole history, past and future, is uncovered.

**PROBLEM 18-2.** Find  $r$  in Eq. (18-2) for a sample of hydrogen gas weighing 2 gm and filling 22.4 liters ( $= 2.24 \times 10^4 \text{ cm}^3$ ) at  $0^\circ\text{C}$  and 1 atm pressure ( $= 1.013 \times 10^6 \text{ dynes/cm}^2$ ). Then determine the volume of this sample at room temperature ( $21^\circ\text{C}$ ) and a gauge pressure of 100 atm. (*Note*:  $0^\circ\text{C}$  and 1 atm = "normal temperature and pressure." Gauge pressure is by definition the pressure *above* atmospheric.)

**18-3 Two gas models.** One is immediately tempted to interpret these observations and the experimental gas law which summarizes them in terms of some model or structure. Even qualitatively, the great compressibility of all gases, the "spring of the air," seems to call for particular explanation. In 1660 Boyle himself, while favoring one model "without peremptorily asserting it," presented two opposing critical explanations by models, both atomistic, for he and his time had begun to accept atomism as part of the rising materialistic world view. Expressed in his words "for all I can clearly discern, whatsoever is performed in the material world, is really done by particular bodies, acting according to the laws of motion."

The two views which Boyle presents may be called respectively the *static* and the *kinetic* views. If air consists of contiguous particles at rest, the corpuscles must themselves be compressible, rather like little springs or, as Torricelli had preferred, like pieces of wool. If, on the other hand, the corpuscles of the gas do not touch at all times, they need be neither variable in size themselves nor maintained by a static assembly, but then they must be in violent agitation, being whirled through all available space within a turbulent though "subtle" fluid. We recall that the subtle fluid theory, then recently proposed by Descartes to account for the motion of planets, had captured the imagination throughout Europe; and indeed, it seemed to explain qualitatively the behavior of gases well enough.\*

Having named the first of these the static model of a gas, and the second the kinetic model, we must make a brief comment on each. As to the first, a springlike compressibility of static particles cannot

---

\* Boyle attempted to find experimentally this subtle fluid, separated from the particles on which it supposedly acted. This was the start of a 250-year futile search for *direct* evidence of ethereal fluids.

easily account for one very conspicuous property of gases—their ability to expand infinitely in all directions, unless of course we allow the improbable, namely, that each atom itself can grow infinitely on expanding. For a pile of ordinary springs opposes compression but does not disperse continuously when put into the open. Therefore some followers of the static view found it necessary to consider that gas corpuscles do not change in size but can repel one another at a distance. Evidently some special new agency of repulsion, some new force among gas atoms, would have to be invented, because the well-known mutual gravitational force between large bodies, the logical candidate for intercorpuscular forces also, unfortunately was always found to be a force of attraction, as is evidently also the nature of those forces which hold together the particles in *solids* and *liquids*.

The distinguished authority of Newton seems to lend some aid to the static picture, for in the *Principia* he had shown, by way of a mathematical demonstration, that *if* a sample of gas were made of mutually repulsive particles, and *if* the force of repulsion between any two particles is *inversely proportional* to the distance between their centers, then the gas pressure in this sample will increase with a small decrease of volume, just as found by Boyle's law. Newton did not mean to prove that gases really do correspond to this model; on the contrary, he writes, "But whether elastic fluids do really consist of particles so repelling each other, is a physical question," i.e., one which has to be settled by appeal to experience. And indeed, this postulated type of force would cause gases to exhibit different behavior for different *shapes* of the container, which does not correspond to anything observed. Nevertheless, Newton's was the first *quantitative* deduction obtained for the behavior of gases on the basis of any atomistic hypothesis, and it appears to have been quite impressive.

The second or kinetic view of the gas might well have been formulated without recourse to an imponderable fluid (if it had not been for the inescapable, almost hypnotic hold of that idea), for it had been the opinion of Greek atomists, as reported by Aristotle and Lucretius, that particles, particularly those not joined in specific form, are "ever passing about in the great void," just as minute specks of dust can be seen dancing irregularly in a sunbeam that pierces a dark room.\*

\* If it be objected that such incessant motion of corpuscles *per se* is impossible, then the eternal turbulence of Descartes' vortices is open to the same objection.

In fact, the Swiss physicist Daniel Bernoulli in 1738 did publish a prophetic, quantitative development of a Lucretian type of kinetic model for gases. Bernoulli thought of the "corpuscles" of the gas as so minute as to be "practically infinite in number" under ordinary conditions, even in a small container. In their rapid motion hither and thither these corpuscles collide with one another and also with the rigid walls of the closed vessel. But the collisions can be assumed to be perfectly elastic; therefore, as we would put it, the kinetic energy of the particles is conserved and the motion can continue undiminished. The pressure which the gas is expected to exert against all sides of the container is thus caused by the incessant impact of millions of high-speed particles—hence the name *impact theory* of gas pressure.

Let us say that the gas-filled container used here is cylindrical, and that the top end can be made to slide in like a piston (see Fig. 18-4). If the volume is slowly decreased, the corpuscles will be more crowded in the progressively smaller space and the number of collisions per second with the walls will be larger, i.e., the pressure should become greater, as actually observed. Bernoulli even calculates the magnitude of this expected increase and finds that it corresponds to Boyle's experimental law. Then he continued:

"The elasticity of air [the pressure in the container] is not only increased by the condensation [decrease of volume] but by heat supplied to it, and since it is admitted that heat may be considered as an increasing internal motion of the particles, it follows that if the elasticity of the air of which the volume does not change is increased [by heating the sample of gas], this indicates a most intense motion in the particles of air; which fits well with our hypothesis . . ."

Lastly, the recognition that  $PV/T$  is constant for a sample of gas is, in a general way, implied in the final summary.

This model, this line of reasoning, and this result, are all quite similar to the work which

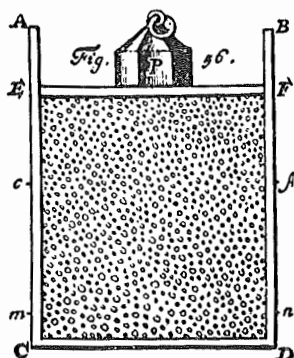


FIG. 18-4. The pressure on the piston exerted by the weight  $P$  is balanced by the impact of high-speed gas particles. (From Daniel Bernoulli's *Hydrodynamica*, 1738.)

more than a century later finally clarified simultaneously the main problems of the nature of gases, heat, and chemistry. But at the time, Bernoulli's work was generally neglected. That lack of general attention, that long interval, are both remarkable, and are not entirely explainable by pointing out that gases were not yet clearly understood by chemists, that density measurements were lacking which would have been needed to obtain the numerical values for the average speed of the particles, and so forth.

Before we turn to ponder the very different view on heat which held the stage for about 100 years after his words were written, we must notice again Bernoulli's phrase ". . . it is admitted that heat may be considered as an increasing internal motion of the particles . . ." Perhaps we may compare Bernoulli's astonishing insight to some quick, isolated, premature stab from a general holding position, far through enemy lines, bringing no strategically valuable consequences until in time the whole main force has moved ahead to link up with this almost forgotten outpost. For Bernoulli in effect had made in his thinking two enormous jumps, for which the temper of the time was not ready for over three generations: first, the equivalence between heat and energy, which implied a general law of conservation of energy, and secondly, the idea that a well-defined numerical relationship, such as Boyle's simple law, could be deduced from the chaotic picture of randomly moving particles.

**18-4 The caloric theory of heat.** Parallel to the story developed to this point there grew up another set of concepts to explain a separate class of phenomena. We recall that Aristotle regarded Fire as one of the four elements, and that by this assumption the observed behavior of heated bodies could indeed be accounted for in a qualitative manner. The ancient Greek atomists explained differences in the temperature of bodies—their so-called intensity or degree of heat—by a closely related conceptual scheme, picturing heat also as a special though not simply and directly perceptible substance, atomic in structure like all the others, quick to diffuse through bodies, and presumably possessing some weight.

As a first step this picture is common-sensical, intuitively clear, even beautiful. Although we may quarrel about details, perhaps about the exact weight of the hypothetical heat substance, the concept on the whole is useful in explaining most of the casual observations, including the temperature equilibrium which two equally hot bodies will eventually attain while in contact with each other. By

and by the picture became sharper: that of a tenuous fluid, free to pass in and out of the smallest "pores" of other materials, present in amounts depending on the temperature, yet otherwise imponderable and perhaps after all rather like Descartes' subtle, omnipresent fluid or the ethers then invoked by some to explain gravitation, the propagation of light and of radiant heat, the transmission of magnetic and electric forces, and the like. But those philosopher-scientists of the 17th century, including Bacon, Galileo, and Boyle, who, when they accepted a fluid theory at all, did so most hesitatingly, felt inclined to distrust the fluid theory of heat altogether and instead, as Bernoulli's excerpt hinted, tended to the general though yet undefined view that heat was directly explainable as a vibration or similar small-scale motion of the atomic particles of heated bodies. The arguments were vague—for example, that heat evidently could be *generated* by motion, as by friction between two sticks or between an axle and the hub of a poorly lubricated wheel.

Throughout the 17th century and far into the 18th, neither the fluid view nor the motion view of heat became firmly established and here the reason was twofold. First, certain key concepts in the study of heat were still missing (e.g., specific heat, latent heat, or even the consistent differentiation between heat and temperature), and second, reliable methods of thermometry and calorimetry were still being developed.\* In the 1780's the great pioneer in the study of heat, Joseph Black of Glasgow and later Edinburgh (1728-1799), still referred uncertainly to heat as "this substance or modification of matter."

But one significant truth had begun to dawn in the first half of the 18th century. In experiments on mixtures, where bodies of varying temperature are brought together, *heat is neither created nor destroyed*; that is, no matter how diverse the redistribution of heat among the different bodies in mixtures or in close contact, the total amount of heat remains constant. In modern terminology we should represent this *law of conservation of heat* in the following manner for the specific and simple case involving two differently hot bodies *A* and *B*, say a hot block of metal (*A*) plunged into a quantity of cool water (*B*): the loss of heat by *A* is equal to the gain of heat by *B*, assuming, of course, that the system containing bodies *A* and *B* is heat-insulated from all outside influences or losses. If *A* and *B*

---

\* For the whole story, well told, see "Case 3, The early developments of the concepts of temperature and heat," by D. Roller, Cambridge: Harvard University Press (1950).

change only their temperatures, but neither their chemical consistencies nor their physical states, the previous sentence can be symbolized by

$$\Delta Q_A + \Delta Q_B = 0,$$

where  $\Delta Q$  is the heat change for any sample, and may be calculated from the defining equation

$$\Delta Q \text{ (cal)} = \text{mass of sample (gm)} \times \text{specific heat} \\ [(\text{cal/gm})/^{\circ}\text{C}] \times [\text{final temperature} - \text{initial temp (in } ^{\circ}\text{C)}].$$

Perhaps this, the basic rule of *calorimetry* ever since, helped most at the time to sway the pendulum of opinion toward the view that heat is a material substance, for such a law of conservation of heat very naturally brought to mind the tacitly accepted concept of the *conservation of matter*. In fact the fluid theory of heat now gave a simple and plausible picture of what happens when differently heated materials are brought together: the excess of "heat fluid" would flow from the hotter to the colder body until an equilibrium had been achieved. And truly, it was almost inconceivable that such a simple observation could be accounted for by the view of heat as a mode of motion of particles.

From the 1760's on, the developments proceeded at an accelerating pace, first mainly through the work of Black and his colleagues. The hypothetical *igneous* or *heat fluid* (named *caloric* by Lavoisier in 1787) began to take on greater plausibility when two further properties were assigned to it: that the particles of the caloric, unlike ordinary matter, repel one another but yet are attracted to the corpuscles of ordinary matter. Therefore, if heat is applied to a material object, the caloric may be pictured as diffusing rapidly throughout the body and clinging in a shell or atmosphere around every one of the corpuscles. If these corpuscles (whose caloric shells repel one another) are free to move apart, as indeed they are in a gas, they will tend to disperse, and more strongly so the greater their crowding or the heat applied, in accordance with the gas laws. If, however, a heated object is in solid or liquid form, the mutual attraction among the corpuscles themselves (considered to be a gravitational one) so predominates that the caloric atmospheres can provide mutual repulsion sufficient only for the well-known slight expansion on heating. Or at any rate, the attraction predominates until enough caloric has been supplied for eventual melting or vaporization.

The same model made it plausible that on sudden changes of the



physical or chemical states the particles of matter will exhibit a sharply changing capacity of accommodating caloric atmospheres, that is, they would release some caloric, thus changing it from a "latent" state to a "sensible" one, where caloric would communicate itself also to neighboring bodies and heat them up. Obviously this would explain the heat generated in some chemical reactions or solutions. Similarly it was not at all unexpected to find that "sensible heat" develops during mechanical deformations: for example, the sometimes very marked rise in temperature of a gas on rapid compression, or of an iron bar by skillful and rapid hammering, was regarded simply as evidence of a squeezing-out of caloric from among the atoms as these are forced closer to one another.

Other experimental facts, including those accessible through the newly developed concepts of specific and latent heat, were subjected to similar and quite successful analyses. The conceptual schemes gradually filled out on all sides, and the gathering weight of evidence persuaded almost all Europe from the 1780's on of the usefulness of the caloric theory.

Yet this theory hid two assumptions, one of which proved fatal to it. The first, which could not by itself discredit the whole structure when brought into the open, was the old problem of whether the caloric material also possesses weight. Naturally, this question was now repeatedly and often ingeniously examined by direct experimentation. But the findings were negative, where they were not indeed contradictory, owing to the inherent difficulties of weighing accurately at varying temperatures.

One most interested and capable experimenter was the amateur scientist of whom we have heard before, Benjamin Thompson, by then Count Rumford (Fig. 18-5). After exhaustive and to us convincing experiments\* with some of the best balances in Europe, he stated in 1799:

"I think we may safely conclude that *all attempts to discover any effect of heat upon the apparent weight of bodies will be fruitless.*"

Granting this conclusion, we discover that we have been brought back to the same ground that more than a century before had been fought over by the vacuists *vs.* the plenists, Newton *vs.* the Cartesians, in a sense even by Galileo *vs.* the Scholastics, namely, the question whether science can accommodate an imponderable fluid with mate-

\* Rumford's experiments, and the discussion of them, make most enjoyable and instructive reading in Roller's account, *op. cit.*, section 3.



FIG. 18-5. Benjamin Thompson, Count Rumford (1753-1814).

rial properties, endowed perhaps even with atomic structure, but inaccessible to direct experimentation or even contemplation through its lack of that most general characteristic of materiality, weight. Those 17th century philosophers who had rejected imponderables—and like them, Rumford—had done so in part because they simply could not imagine a weightless fluid, a mechanical entity lacking one of the most unforgettable properties of ordinary matter. But calorists, on the other hand, could equally well say with Black: “I myself cannot form a conception of this eternal tremor [motion of corpuscles] which has any tendency to explain even the more simple effects of heat.” Up to this point, Rumford may seem to have little cause for rejecting the imponderable calorics except a lack of imagination; and this is the more striking because the scientific world at the time generally did accept a whole set of imponderables and ethers.

Let us for a moment examine our own first impulses in this matter. We are probably tempted to reject the concept of imponderable fluids outright; and if urged to explain, we might say that a part of the very definition of the physical entity must be that we can discover it directly. But this would be a most naive attitude; we must discard it now before we go on to phenomena involving “imponderables” like electrons and photons. We must be prepared to accept not just what sounds plausible, or just what can be imagined by analogy (as the atomists accepted smaller corpuscles by analogy

with larger bodies), but we must be prepared to welcome into science any concept that may be needed in our schemes if only it fulfills these requirements: it should be definable, without internal contradiction and ambiguities, in terms of observable and measurable effects whether these be direct or not; and it should prove useful in a wide variety of descriptions and laws, long known as well as newly deduced. The caloric fluid fulfilled all of these criteria at the time.

But this was not the reasoning which led the calorists to accept the imponderable, and Rumford was not just naive in rejecting it. They on the one side were not being greatly bothered by Rumford's demonstrations, and on the other side his exception to the caloric for lacking measurable weight was only part of his much more thorough experimental, thirty-year attack on that concept. In time, Rumford succeeded in discovering the real Achilles heel—a failure of that most treasured function and main justification of the concept of the caloric, namely, *its usefulness in the general law of conservation of heat*.

The famous experiments leading up to this far-reaching discovery are described by him in 1798 as follows:

“[I] am persuaded that a habit of keeping the eyes opened to everything that is going on in the ordinary course of the business of life has oftener led—as if it were by accident or in the playful excursions of the imagination, put into action by contemplating the most common occurrences—to useful doubts and sensible schemes for investigation and improvement than all the more intensive meditations of philosophers in the hours especially set apart for study.

“It was by accident that I was led to make the experiment of which I am about to give account . . . Being engaged lately in superintending the boring of cannons in the workshops of the military arsenal at Munich, I was struck by the considerable degree of heat [temperature] that a brass gun acquires in a short time in being bored, and with the still higher temperature (much higher than that of boiling water, as I found by experiment) of the metallic chips separated from it by the borer.”

A commonplace type of observation, made accidentally, but heightened both by the uncommon temperature developed and by the exceptional gift of observation of Rumford, was therefore the effective beginning of the idea that heat *fluid* in enormous quantities can be generated by friction. As we saw, that of itself is not al-

together unexpected; on the basis of the caloric theory the mechanical treatment of metals in boring might decrease the capacity for holding caloric in the metal, thereby yielding free, "sensible" heat. However, experiments with the metallic chips, and with the bulk metal of the cannons themselves, showed Rumford that their capacity for heat (specific heat) had not changed. What is much more significant, further experiments revealed that the source of heat generated "in these experiments appear to be inexhaustible. It is hardly necessary to add that anything which any *insulated* body, or system of bodies, can continue to furnish *without limitation*, cannot possibly be a *material substance*, and it appears to me extremely difficult, if not quite impossible, to form any distinct idea of anything capable of being excited or communicated in the manner in which heat was excited and communicated in the experiments, except it be MOTION." And throughout his voluminous writing Rumford continues to affirm those "very old doctrines which rest on the supposition that heat is nothing but a vibratory motion taking place among the particles of the body."

A similar conclusion was drawn the following year by the 21-year-old Humphry Davy, later a distinguished chemist, in original experiments in which he allowed two pieces of ice to rub against each other until they had both melted. The large amount of heat needed for melting the ice must have been produced by friction only, particularly in view of the large increase of specific heat on changing ice into water. Davy's conclusion that "it has just been experimentally demonstrated that caloric, or matter of heat, does not exist," although very influential on the later work of Mayer and Joule, is rather an overstatement; he might have been satisfied with the claim that large quantities of caloric appeared to have been created by friction, contrary to the expectations of the calorists.\*

In brief, these and a great variety of other experiments, executed mostly by the indefatigable Rumford, had struck key blows against the caloric theory by 1807. But that was not enough to convert the calorists and topple their science, for the simple reason that there was no good constructive conceptual scheme to offer instead. Even though Rumford was able to show that the kinetic model would account qualitatively quite well for a variety of observations such as normal thermal expansion, heat conduction, change of state, and

---

\* There is an ironic twist here. Re-examination of Davy's data and methods has shown that the ice must have melted owing far more to heat leaking into his apparatus than to friction: see Roller, *op. cit.*, pp. 86-87.

diffusion in liquids, his picture of heat as a mode of motion was in general even more qualitative than Bernoulli's; it could not explain the undisputed facts summarized by the law of conservation of heat in *mixtures* and in other such simple cases where, as we now would put it, by careful insulation the amount of heat in the system is kept constant, and is neither generated nor converted. The caloric theory was still most practical and plausible there and in a great region of other physical and chemical phenomena. We see here a parallel to the assault in Boyle's writing on chemistry: in trying to reformulate chemistry he, too, had succeeded only in throwing doubt on the old system, weakening it, and preparing the way for the eventual overthrow later by a more complete theory.

What finally had to be done here (by Mayer and Joule in the 1840's) was to show that heat was interconvertible with mechanical work according to a constant rate of exchange, and that the particle picture was, in quantitative predictions, superior to the caloric scheme. At any rate, in Rumford's and Davy's work there were two related conclusions to team up with those reached before by Newton, Hooke, Bernoulli, and other atomists: that it may be thought that matter is made of discrete particles, and that heat is to be explained not by a separate substance but in terms of the behavior of those matter-particles.

Nothing else of importance or decisiveness was added to this controversy over the nature of heat until over half a century after Rumford's experiments. It was not a matter of looking for crucial experiments to decide between the kinetic theory and the caloric theory; Rumford's experiments on the unlimited creation of heat by friction would perfectly have served this purpose if the calorists had been ready for a change of heart. In actual fact, the next step of the development, as so frequently happens, came from a completely unexpected direction: through the atomic theory of chemistry by John Dalton. As we turn to his work we shall watch the following sequence: Dalton, a firm calorist, first seems to establish the caloric theory of heat and the atomic theory of matter side by side. In time there are seen to develop contradictions within his scheme which tend to discredit his whole work. Then, through the contributions of Avogadro and Cannizzaro, Dalton's concepts are amended, essentially by abandoning that part of the conceptual scheme which previously depended on the properties of the caloric, and by assigning to the atom itself the dual task of accounting for both the physical and the chemical behavior.

PROBLEM 18-3. Was the concept "caloric fluid" right? Was it wrong? Why? Analyze the caloric theory in terms of the criteria for a good theory (Section 8-2).

#### Further References

J. B. Conant, *op. cit.* (Ch. 15). See pp. 63-108 and 142-152 on the concept of atmospheric pressure and Boyle's work.

J. B. Conant, *Robert Boyle's Experiments in Pneumatics*. Cambridge, Mass.: Harvard University Press (1950).

J. C. Gregory, *A Short History of Atomism*. London: A. and C. Black (1931), Chapters I-VI.

P. Lenard, *op. cit.* (Ch. 9); biographies of Boyle, Rumford, Davy.

W. F. Magie, *op. cit.* (Ch. 2). Excerpts from Torricelli, Boyle, Bernoulli, Black, Rumford, and Davy. Also W. S. Knickerbocker, *op. cit.*, Chapters VI and XXI; and H. Boynton, *op. cit.*, pp. 173-198, 225-229, 242-247.

D. Roller, *The Early Development of the Concepts of Temperature and Heat: The Rise and Decline of the Caloric Theory*. Cambridge, Mass.: Harvard University Press (1950).

## CHAPTER 19

### THE ATOMIC THEORY OF CHEMISTRY

**19-1 Introduction.** When in the preceding chapter we used the words *atom*, *corpuscle*, *particle*, and the like, it was not to be inferred that the writers of antiquity or even of the 17th and 18th centuries had in mind what we mean by atom. The atomistic idea at that time served the purpose of providing logical satisfaction for some of the conceptual schemes in physics and chemistry, usually mechanical models which postulated small indivisible bodies by which certain of the constantly recurring or persisting properties could be explained.

As we have seen, the atomistic hypotheses up to 1807 had not given rise to an unambiguous and precise theory either of matter or of heat. It may be said that the presumption for atoms was large but the choice for particular details wide. Yet, between the publication of Robert Boyle's *Sceptical Chymist* in 1661 and Lavoisier's death in 1794, there had developed a fundamental concept which was destined to give a new, vital, and precise meaning to the speculative atomistic hypothesis, namely, the concept *chemical element*. Boyle had tried to wrest it from the alchemists of Aristotelian and Paracelsean origin, and had proposed that chemistry must be built on an identification and knowledge of those chemical substances which cannot be separated into different components by any common means. Only those substances, he maintained, were worthy of the suggestive and ancient name "element."\* By the end of the 18th century this idea had indeed finally proved to be the sound basis for explaining such chemical reactions as combustion.

Lavoisier himself, in the preface to the *Elements of Chemistry* (1789), had written this revealing and, in part, quite modern analysis of the concept "element" and its relation to the atomistic hypothesis of his predecessors:

---

\* "And to prevent mistakes I must advertise you, that I now mean by elements, as these chymists that speak plainest do by their principles, certain primitive and simple, or perfectly unmingled bodies; which not being made of any other bodies, or of one another, are the ingredients of which all those called perfectly mixed bodies are immediately compounded, and into which they are ultimately resolved."

“All that can be said upon the number and nature of elements is, in my opinion, confined to discussions entirely of a metaphysical nature. The subject only furnishes us with indefinite problems, which may be solved in a thousand different ways, not one of which, in all probability, is consistent with nature. I shall therefore only add upon this subject, that if, by the term *elements*, we mean to express those simple and indivisible atoms of which matter is composed, it is extremely probable we know nothing at all about them; but, if we apply the term *elements*, or *principles of bodies*, to express our idea of the last point which [chemical] analysis is capable of reaching, we must admit, as elements, all the substances into which we are capable, by any means, to reduce bodies by decomposition. Not that we are entitled to affirm, that these substances we consider as simple may not be compounded of two, or even of a greater number of principles; but, since these principles cannot be separated, or rather since we have not hitherto discovered the means of separating them, they act with regard to us as simple substances, and we ought never to suppose them compounded until experiment and observation has proved them to be so.”

In addition to the few elements known since the ancients (e.g., metals, sulfur), a growing number of new ones had been identified (about one-third of our present list) although, of course, many recalcitrant compounds still posed as elements until electrolysis and other powerful means became available to separate their components. Not so very surprisingly, the caloric also was generally listed among the elements.

The combination of the ideas “chemical element” and “atom” is now in retrospect perhaps all too obvious. An atom of a chemical element would be the smallest physical particle that can participate in the identifying chemical reactions on its own. The thought urges itself on our minds to postulate, until evidence to the contrary appears, that each atom of an element is identical to every atom of that same element, although we may be prepared to allow that different elements have physically quite different atoms. An element in the gaseous state, say oxygen or hydrogen, should surely consist of individual atoms, and a chemical compound, say water vapor, might well be imagined as constituted of particles, each containing two or more atoms of different elements. In particular, it being only reasonable to again make initially the simplest assumption, the smallest corpuscles of water vapor (we might call them *atoms*



of water) should be regarded as made of a close union of one atom each of hydrogen and oxygen, the two gases which would yield water when mixed and exploded.

This, in brief, introduces us to the thoughts of the Englishman, John Dalton (1766-1844), who in his work from 1800 on effectively performed that marriage of the idea of chemical elements and the ancient atomistic hypothesis. Dalton was largely self-taught. A retiring person, son of a humble hand-loom weaver, he supported himself poorly as teacher (as early as his twelfth year) and general tutor in Manchester. While possessing a strong drive and a rich imagination, particularly along the line of mechanical models and clear mental pictures, his supreme gift was his astonishing physical intuition, which allowed him to arrive at correct conclusions despite being only "a coarse experimenter," as Davy called him.\*

**19-2 Dalton's model of gases.** Like Kepler and Newton, whom he resembled in many ways, Dalton published on a great variety of subjects; but it appears that his main work originated in an interest in the *physical* structure of gases. He reports that through his (competent) interest in meteorology he had been led to wonder why the earth's atmosphere, being composed of a mixture of several gases of different density, e.g., nitrogen, oxygen, and water vapor, should be so homogeneous, that is, why samples taken at not too widely differing altitudes should have the same proportion of chemical elements, even though nitrogen gas, being lighter per unit volume in the pure state than oxygen gas, might be expected to "float" on top of the latter, somewhat like oil on water. Rumford would have had a good answer; he might have said, as he did in an analogous case of mixed liquids, that by thermal agitation, by their "peculiar and continual motion," all the individual particles must diffuse among one another and thus mix thoroughly. However, Dalton could never accept this kinetic view of matter for a moment. He adopted the static model of a gas, being (wrongly) convinced that the great Newton had *proved* it in the brief passage from the *Principia* to which we have referred before. "Newton had demonstrated," Dalton once wrote, "from the phenomena of condensation and rare-

---

\* For an account of Dalton's work which is as highly recommendable as it has been indispensable for the preparation of this chapter, see Leonard K. Nash, *The Atomic Molecular Theory*, Case 4 in the Harvard Case Histories in Experimental Science; Harvard University Press (1950). See also Sir Henry E. Roscoe and Arthur Harden, *A New View of the Origin of Dalton's Atomic Theory*, Macmillan (1896).

faction that elastic fluids are constituted of particles, which repel one another by forces which increase in proportion as the distance of their centers diminishes: in other words, the forces are reciprocally as the distances. This deduction will stand as long as the Laws of elastic fluids continue to be what they are."

Dalton was naturally forced to reject the kinetic view of heat, along with the kinetic view of gases, and to adopt the caloric theory which was then, of course, still the general conceptual scheme for heat phenomena. The only satisfactory solution to the problem of homogeneity of gas mixtures based on a static model was, to Dalton's mind, the following set of assumptions and deductions, some of them taken from the work of predecessors:

(a) *Each particle of a gas is surrounded by an atmosphere of caloric.* As Dalton put it in his main work, "A vessel full of any pure elastic fluid presents to the imagination a picture like one full of small shot. The globules are all of the same size; but the particles of the fluid differ from those of the shot, in that they are constituted of an exceedingly small central atom of solid matter, which is surrounded by an atmosphere of heat, of greater density next the atom, but gradually growing rarer according to some power of the distance."

(b) *The particles are essentially at rest, their shells of caloric touching.* At the top of Fig. 19-1, taken from Dalton's notebook, is shown the case of two neighboring gas atoms. Throughout, Dalton accepts the notion that the only means for the communication of forces among atoms is direct contact. He is as uncomfortable with the notion of action at a distance as had been Newton's contemporaries.

(c) *The total diameter of each particle, counting in the caloric atmosphere or shell, differs from one substance to the next.* ". . . the following may be adopted as a maxim, till some reason appears to the contrary: namely,—That every species of pure elastic fluid has

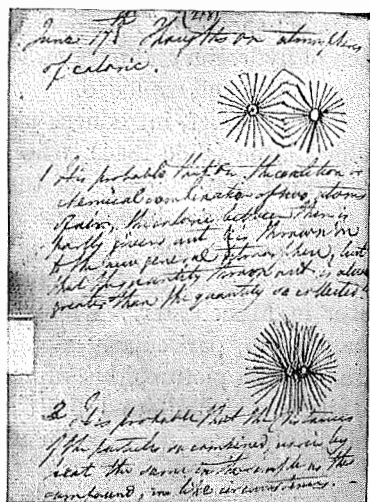


FIG. 19-1. From Dalton's Notebook. At the top: two elementary atoms, as they would exist in a gaseous mixture. The straight lines indicate the region of caloric around each central mass. Below: a molecule of a compound, consisting of two atoms in close contact, surrounded by a common, nearly spherical atmosphere of caloric. (As reproduced by Roscoe and Harden.)

its particles globular and all of a size; but that no two species agree in the size of their particles, the pressure and temperature being the same."

This point proved crucially important in the later development. Dalton thought he had deduced this maxim unambiguously from experimental evidences of the following type. One of the products formed when oxygen and nitrogen gas are combined is nitric oxide or, as Dalton called it, "nitrous gas." On other grounds, Dalton had decided that in this gas one nitrogen atom combines always with just one oxygen atom. However, in his own experiments, it seemed that the *ratio by volume* of oxygen gas to nitrogen gas needed in the reaction was not 1:1, which would have indicated equal volumes for both types of atoms, but more nearly about 0.8:1, pointing to the conclusion that nitrogen atoms were somewhat larger than oxygen atoms. (In actual fact, the experimental results here are easily misleading because several different compounds of nitrogen and oxygen may be formed concurrently.)

The same set of experiments seemed to furnish Dalton with another confirmation of his "maxim." The nitrous gas so formed took about double the volume of either of the component elements. In terms of a gas whose particles all touch, this seemed to indicate clearly that the particles or (compound) "atoms" of nitrous gas take about twice the volume of either oxygen or nitrogen particles.

Experiments on water vapor seemed to bear out the same type of conclusion: "When two measures of hydrogen and one of oxygen gas are mixed, and fired by the electric spark, the whole is converted into steam, and if the pressure be great, this steam becomes water. It is most probable then that there is the same number of particles in two measures of hydrogen as in one of oxygen . . ."—that is, hydrogen atoms are about twice as large as those of oxygen.

(d) Turning back to the initial problem, Dalton concluded on the basis of some qualitative speculations that there could be no stratification of the air into its elements *because the contiguous particles of several different sizes would tend to push one another away until stability is reached in a homogeneous mixture.*

(e) Therefore, he reasoned, the constituents of the atmosphere must have been mixed with one another in the disequilibrium during their original encounter, and since then have remained homogeneous.

This suggestive "pile of shot" picture of the atmosphere prompted Dalton to consider next the number and relative weights of atoms, particularly "of all chemical elementary principles which enter into any sort of combination one with another." But before we turn to

look at the enormously fruitful results that followed this decision, let us in passing note the amazing fact that each and every one of the points (a) to (e) above is, from the present point of view and often even in terms of internal consistency, *wrong*; that furthermore the premise itself, Newton's supposed proof of the static model, is Dalton's misinterpretation. Are we to conclude that science is a blind man's game, that we must grope from error to error, making a few valid discoveries only by chance, by sheer luck, or simply as by-products of a prodigious amount of largely fruitless labor being done everywhere and throughout the centuries? This is a good occasion for introspective reflection. The answers to our questions must surely all be negative, if only for the specific reasons that out of Dalton's own work, for example, grew chemistry as we know it.

**19-3 The properties of Dalton's chemical atom.** But to return to the growth of the atomic theory of chemistry: Dalton's great opus, *A New System of Chemical Philosophy*, from which most of the previous quotations have been taken, was first published in two parts, in 1808 and 1810. From that work and from other writings, we can abstract the several principles which reflect the basic ideas in his conceptual scheme:

(a) *The existence of indivisible atoms*: "Matter, though divisible in an *extreme degree*, is nevertheless not *infinitely* divisible. That is, there must be some point beyond which we cannot go in the division of matter. The existence of these ultimate particles of matter can scarcely be doubted, though they are probably much too small ever to be exhibited by microscopic improvements. I have chosen the word *atom* to signify these ultimate particles . . ."

(b) *Atoms are unchangeable*: The atoms of different elements "never can be metamorphosed, one into another, by any power we can control," as the failure of centuries of alchemy had made clear. By continual *test* the transmutation of elements was found to be impossible—a *postulate of impotency* arrived at in a manner exactly analogous to the rise of the law of conservation of energy from the long failure to find a perpetuum mobile.

(c) *The structure of compounds*: Chemical compounds are formed by the combination of atoms of two or more elements into a "compound atom," although we now name the smallest particle of a compound a *molecule*. "I call an ultimate particle of *carbonic acid* a *compound atom*. Now, though this atom may be divided, yet it ceases to be carbonic acid, being resolved by such division into char-

coal and oxygen. Hence I conceive there is no inconsistency in speaking of compound atoms, and that my meaning cannot be misunderstood."

Although we may be tempted to think of the particles of compounds as made of two or more atoms, so to speak, side by side, and although Dalton's ideographs, as we shall see at once, at first glance give the same impression, he did not have this conception. Rather, he thought that the compound molecule was essentially round or "globular," the centers of the combining atoms being "retained in physical contact by a strong affinity, and supposed to be surrounded by a common atmosphere of heat." (See Fig. 19-1.) In the specific example of the formation of water, he asked that it should "be supposed that each particle of hydrogen attaches itself to a particle of oxygen, and the two particles so united form *one*, from which the repulsive energy emanates [by a redistribution of the combined caloric]: then the new elastic fluid may perfectly conform to Newton's Law . . ."

(d) *The identity of all atoms of one substance*: "the ultimate particles of all homogeneous bodies are perfectly alike in weight, figure, etc. In other words, every particle of hydrogen is like every other particle of hydrogen . . ." We should here realize that Dalton was now simplifying even beyond the older atomists, who usually had felt it necessary to allow different sized atoms for the same element, just as there might be found different pebbles all made of the same material.

(e) *The nature of chemical reactions, and the conservation of mass*: "Chemical analysis and synthesis go no farther than to the separation of particles one from another, and to their reunion. No new creation or destruction of matter is within the reach of chemical agency. We might as well attempt to introduce a new planet into the solar system, or to annihilate one already in existence, as to create or destroy a particle of hydrogen. All the changes we can produce, consist in separating particles that are in a state of cohesion or combination, and joining those that were previously at a distance."

We have already noted a passage in which the combination of hydrogen and oxygen into water vapor is pictured. To represent to himself this and other chemical reactions, Dalton found it extremely helpful to make use of simple but ingenious symbols (Fig. 19-2). Even the earliest alchemists had represented different substances by different symbols, but Dalton's ideographs represented *individual atoms*, something that had not been done before. But note well that in our table some of the first twenty symbols do not represent

the atoms of elements but the molecules of compounds. For example, symbols 9 and 10, which stand for soda and potash, were known to Dalton from Davy's work of the same year (1808) to be "metallic oxides" (actually metallic hydroxides).

Because we shall make some use of these historic symbols, let us briefly note those for the compounds. For example, 21 refers to a binary compound made of one atom of oxygen and one of hydrogen, namely water; 22, "an atom of ammonia, composed of one of azote [the then current name for nitrogen] and one of hydrogen." The 23rd represents "an atom of nitrous gas [now called nitric oxide] composed of one of azote and one of oxygen"; 24 was Dalton's symbolism for what we now call methane, and 25 for carbon monoxide, then called "carbonic oxide." Among the ternary compounds 26 is nitrous oxide, still called thus, although the name laughing gas is more familiar; 27 is a compound atom of "nitric acid," which we would name a molecule of nitrogen dioxide; 28 represents "carbonic acid," now called carbon dioxide.

In the same fashion, Dalton pictured to himself other compounds which he knew to contain more than three elementary atoms, namely quaternary, quintary, etc. Nowadays, of course, we are used to other symbols, those introduced by the Swedish chemist Berzelius in 1819. Letters are used to represent the substance, e.g., hydrogen, oxygen, nitrogen, carbon, and chlorine are respectively H, O, N, C, Cl (the full list appears in Appendix III), and compounds are rendered by the joining of the respective symbols of elements; for example, hydrochloric acid would be represented by HCl, and carbon dioxide by CO<sub>2</sub>. A chemical reaction, say the combustion of carbon (the formation of carbon dioxide), might then be given in these ways,

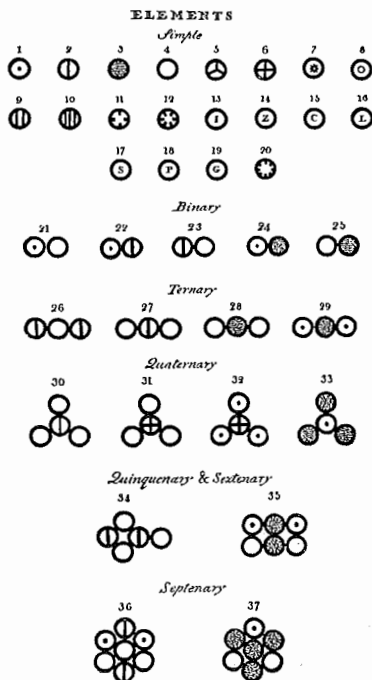
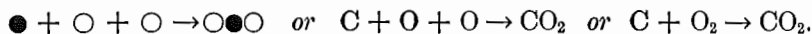


FIG. 19-2. Dalton's symbols for "elementary" and "compound" atoms, taken from his work, *A New System of Chemical Philosophy*, Part I. Dalton writes, "This plate contains the arbitrary marks or signs chosen to represent the several chemical elements or ultimate particles."

the first being Dalton's, the last being the present convention:



(f) In any chemical reaction, *atoms combine in simple, definite ratios* to form compound atoms [molecules]. This part of Dalton's hypothesis, which had been evidently implied in the preceding discussion, denied a then widespread and distinguished opinion concerning chemical reactions, namely, that particularly in the case of two gases the compound is formed by a solution of one in the other. Dalton, a little querulously, gives several arguments "advanced to shew the absurdity of this doctrine . . ." Instead of regarding water as made of an indefinite or variable combination of hydrogen and oxygen particles, Dalton conceived of each molecule of water vapor as made the same way, namely, of one atom of each element. But there's the rub; on the basis of what has developed so far, Dalton has, of course, no way of knowing *just how many* atoms of each element do in fact combine in one water molecule; it might well be two of the one and one of the other. The relative volumes of combining gases do not offer a clue, since Dalton had already assumed the atoms to have different and, of course, unknown sizes. The real evidence for this sixth assumption came not from the somewhat dubious earlier experimental results on the combining volumes of gases, but from the quite firmly established and simple rule concerning *combining weights*. The accumulation of gradually more and more accurate analytic data during the prior half century, and particularly the recent and vigorous dispute between the French chemists Proust and Berthollet, had made increasingly acceptable what we now call the *law of definite proportions*: no matter in which locality a pure substance—say a metallic oxide, a drop of some liquid, or a complex gas—is found, no matter whether it is artificially prepared or taken from nature, the proportions by weight in the constituent elements in every case are always the same. For water, this ratio by weight between hydrogen and oxygen gas is 1:8 (approximate modern value; Dalton's was 1:6 and later 1:7).\* Now, combining the simple pictorial view of

\* We must stop a moment to realize that like the law of conservation of mass, which was the fruit of a fierce struggle between Lavoisier and his followers in their fight against the phlogistonists, so also was the establishment of this law much speeded by the vigor of the dispute between able antagonists. Thereby were also established at the time several other most valuable results, e.g., clearer distinctions between compounds and mixtures, and a large amount of data that provided the raw material for further theorizing. From our dynamic view of science we can well understand this new illustration of the evolutionary mechanism in the struggle of ideas.

atoms and the law of definite proportions of combining weights, Dalton was almost forced to the hypothesis that the participating atoms combined in definite ratios by numbers. However, the assumption that each molecule contains specific and fixed numbers of atoms leaves unanswered the urgent question of what these numbers are. All that can be said is the following, again taking the case of water as an example: the relative combining weights between hydrogen and oxygen gas being 1 to 8, and *assuming* that the combination proceeds by joining a single atom of each *for every single molecule*, then we may conclude that the relative atomic weights of hydrogen and oxygen are as 1 to 8 also. On the other hand, if the molecular formula is not HO but H<sub>2</sub>O (i.e.,  $H + H + O \rightarrow H_2O$  or  $H_2 + O \rightarrow H_2O$ ), then the relative atomic weights would be 1 to 16.

The missing link to an understanding of chemical reactions, therefore, is the knowledge of some simple molecular formulas. There Dalton is finally driven to an *assumption*, one which stands as typical of the advantages and dangers of any hypothesis.

(g) *The rule of greatest simplicity of the most common stable compounds*: "When only one combination of two bodies [elements] can be obtained, it must be assumed to be a binary one, unless some other cause appears to the contrary . . . When two combinations are observed, they must be presumed to be a binary and a ternary . . ." and so forth for more than two combinations. Specifically *water*, the only known combination of hydrogen and oxygen, must be a binary compound (HO or  $\odot\odot$ ).\* In the same manner, the only two known combinations of carbon and oxygen were identified as follows: the lighter of the gases ("carbonic oxide") as CO, the heavier ("carbonic acid") as CO<sub>2</sub>. The common compound of nitrogen and hydrogen, ammonia, was regarded as NH (we now know it to be NH<sub>3</sub>). Ethylene and methane were identified as CH and CH<sub>2</sub> (whereas now they would be written C<sub>2</sub>H<sub>4</sub> and CH<sub>4</sub>).

One brief look at any modern table of chemical formulas will at once convince us of the fallaciousness of this maxim. For example, the only iron-iodine compound, iron iodide, is not binary but has the formula FeI<sub>2</sub>, and the simplest iron chloride is not FeCl but FeCl<sub>2</sub>. The iron phosphides are not FeP, FeP<sub>2</sub>, and Fe<sub>2</sub>P, as Dalton might have expected, but FeP, Fe<sub>2</sub>P, and Fe<sub>3</sub>P. The common oxides of iron are now recognized to be FeO, Fe<sub>2</sub>O<sub>3</sub>, and Fe<sub>3</sub>O<sub>4</sub>. However, the analytic data in most cases (we might almost say luckily) were too ambiguous to point out a large-scale failure of his axiom, and

---

\* Hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) had not yet been identified.

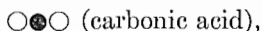


in some cases it gave correct results (e.g., for the carbon oxides and for the nitrogen oxides). Above all, Dalton's devotion to the time-honored idea of "the regularity and simplicity" in the laws of nature did make possible a start, and a fruitful one, toward the identification of molecular formulas. Where some early guess of a molecular formula was wrong it could be corrected through internal consistency checks; we recall again the admonition that through sincere and hard work, truth can come out of error more easily than out of confusion.

**19-4 The early achievements of Dalton's theory.** We might discuss a variety of subtle consequences of Dalton's work: that his pictorial conceptualizations and his ideographs must have produced a deeper feeling for the reality of atomic processes at the time; that his identification of chemical atoms helped in the final differentiation between the concepts of pure chemical compounds and mixtures; and other details of this sort. But above all, we must now summarize briefly the initial main achievements of Dalton's theory—its power to gather known facts and laws, and to provide a simple, workable conceptual scheme; its fruitfulness in suggesting new relations and stimulating research; and its helpfulness in the solution of problems in contemporary science. In the following list, elements of each of these functions may be found in all items.

(a) Dalton's chemical atom is the concept that integrates the law of conservation of mass and the law of definite proportions. The latter, although essentially established at the time by empirical and exhaustive test, was not yet explicitly and generally appreciated. This experimental result was now given physical meaning and supported by the inherent plausibility of Dalton's picture, and in this sense we may say that Dalton did for the embattled law of definite proportions what Newton did for Kepler's established laws. Other established chemical results could be similarly accommodated (e.g., the law of equivalent proportions).

(b) But the best support of the atomic theory, by common consent at the time, was an extraordinarily powerful insight that followed directly and naturally from Dalton's conceptual scheme: when one substance unites with another in more than one proportion, these different proportions bear a simple ratio to one another. This is *the law of multiple proportion*; we see at once that by necessity it is fully implied in the last two of Dalton's assumptions. For example, by picturing "carbonic oxide" and "carbonic acid" in Dalton's manner,



it is clear that we should find associated with a given quantity of carbon twice as much oxygen by weight in the second case compared with the first; or, to put it into the language of the law, the respective weight proportion of oxygen to carbon in the second substance, *divided* by the proportion of oxygen to carbon in the first, should be as 2:1, a simple ratio indeed. As to the experimental check of this prediction, the necessary data for combining weights had, significantly, been long available in about the following form:

In "carbonic oxide," 44% carbon and 56% oxygen (by weight);  
 in "carbonic acid," 28% carbon and 72% oxygen.

If we reduce these approximate percentages to *proportions*, we obtain

In "carbonic acid," the weight of oxygen to carbon  $\doteq$  5:2;  
 in "carbonic oxide," the weight of oxygen to carbon  $\doteq$  5:4.

Therefore the ratio between these proportions =  $(5/2)/(5/4) = 2$  (or 2:1). Similar confirmations of the law of multiple proportions could be obtained by an analogous treatment of other data available at the time. We might say that the facts had been before everyone's eyes, but the simplicity they contained remained hidden because there was previously no good reason to look for it. As has often been said on these pages, one cannot see much without the aid of a hypothesis to organize the perceptions; in a sense, the old saying "seeing is believing" should be supplemented by another, "believing is seeing."

It should be evident that this discovery or prediction of the law of multiple proportions stimulated a great deal of new research. Given any one combining proportion for two substances, one could now plan for and predict the analysis or synthesis of other compounds of the same elements. The results were impressive, although we now know that the postulate of "simple ratios" is again too limited and that very large whole numbers may be involved. For example, in the complex structure of the organic compounds, then still largely inaccessible, each molecule may contain several dozens of atoms of a few substances, so that the ratio of carbon-to-hydrogen proportions is not such a simple one for two substances like  $C_{23}H_{22}O_6$  and  $C_{22}H_{42}O_3$ .

**PROBLEM 19-1.** In 1810, Dalton gave the approximate results listed in the following table for the relative densities of five gases (his terminology), each made of molecules involving only nitrogen and oxygen. He also concluded that the lightest of these gases was made of molecules containing one atom each of N and O. (a) Decide on the molecular formula for each gas on the basis of this information. (b) Confirm the law of multiple proportions for these five gases.

<i>Substance</i>	<i>Relative Density</i>	<i>% by weight</i>	
		<i>Nitrogen</i>	<i>Oxygen</i>
Nitrous gas	12.1	42.1	57.9
Nitrous oxide	17.2	59.3	40.7
Nitric acid	19.1	26.7	73.3
Oxynitric acid	26.1	19.5	80.5
Nitrous acid	31.2	32.7	67.3

(c) A third achievement of the atomic theory of chemistry, one without which modern physics and chemistry would be unimaginable, was the start toward the determination of atomic weights. In Dalton's words, "Now it is one great object of this work, to shew the importance and advantage of ascertaining the relative weights of the ultimate particles, both of simple and compound bodies . . ."—until then, indeed, neither possible nor even meaningful.

It was quite evidently hopeless to expect a determination of absolute atomic weights at that time. We now know, by indirect evidences, that the atoms of elements are of the order of  $10^{-24}$  gm each. But the foregoing discussion has hinted strongly how the *relative* weights of atoms were to be obtained: from the *known combining weights* of the elements of a substance, and from the *assumed molecular formulas*. For example, from problem 19-1 it would appear, as you may check in five ways, that the relative weights of the atoms of nitrogen and oxygen are nearly 6 to 8.2 (from better experiments we now have the values of 7:8, or more precisely, 14.008 to 16.000). We may use other analytic data in the very same way: the combining weights of hydrogen and oxygen in Dalton's synthesis of water being as 1 to 6 (later 1 to 7, and by present measurements nearly 1 to 8), and the formula for water having been assumed to be HO ( $\odot\odot$ ), the relative atomic weights were also at first supposed to be as 1 to 6. By making a compromise between the various results of this type, Dalton could assemble a table of most probable values of relative atomic weights (although the words "relative combining masses" would have been more appropriate). By referring all values to the arbitrary value of unity for hydrogen, Dalton assembled tables for most elements and compounds then available. One of the earlier tables,\* published in 1808, gave the following list of relative atomic weights: hydrogen 1, nitrogen 5, carbon 5, oxygen 7, phosphorus 9,

\* As a glance at the modern table of elements will show, our present arbitrary reference substance for atomic weights is not hydrogen but oxygen, taken to have an atomic weight of 16.0000. On that basis the atomic weight of hydrogen is about 1.008.

sulfur 13, soda (a "compound atom," namely NaOH, sodium hydroxide) 28, and so forth.

### 19-5 Gay-Lussac's law of combining volumes of reacting gases.

It soon became obvious to Dalton that in using the rule of greatest simplicity, inconsistencies were bound to occur in the determination of atomic weights based on various compounds. In the specific case of water he said as early as 1810, "After all, it must be allowed to be possible that water may be a ternary compound" ( $\text{H}_2\text{O}$  or  $\text{HO}_2$  instead of  $\text{HO}$ ). In the former case the (modern) combining weights of 1:8 for the two gases would indicate that the relative atomic weight of oxygen is 16 if hydrogen is 1; in the latter case, if  $\text{H} = 1$ ,  $\text{O} = 4$ . It is conceivable that by allowing such flexibility in the assumption of molecular formulas, and by eliminating contradictions, one might eventually have arrived at a set of consistent atomic weights (the process is illustrated in the next problem). But just then there appeared from France an experimental discovery and from Italy a theoretical insight which at once aided this task, and which together were destined to replace the gradually less and less defensible Daltonian rule of greatest simplicity, putting atomic chemistry on a new basis.

**PROBLEM 19-2.** In the table below is given a set of approximate *combining weights* for five substances. For each, Dalton's principle of greatest simplicity would demand a *binary* structure, although in fact most of them are more complex. Analyze these data and obtain, without recourse to any other aids and exclusively from the consistency of these data, (a) a set of the most consistent molecular formulas for all these substances, and (b) a set of the most consistent atomic weights for all the elements involved.

Then consult a collection of constants for inorganic compounds (as for example in the *Handbook of Chemistry and Physics*) and find the correct formulas and the current names for these compounds.

Substance	Combining weights in gm			
	Iron	Sulfur	Hydrogen	Carbon
A	14			9
B		16	1	
C	7	8		
D		16		3
E		48	1	6

(Hint: Begin by writing for each of the *first four* substances several possible molecular formulas, and the relative atomic weights for each of these

cases as they would follow from the given combining weights. Next, select the one most consistent set of atomic weights, and thereby you have decided on the most probable molecular formulas also. *Then*, as a test of your scheme, turn to the last substance and find whether it can be accommodated.)

In 1808 Joseph Louis Gay-Lussac (see Section 18-2), a master of experimental chemistry, had noted that Cavendish's experiment on the explosion of hydrogen and oxygen gas to yield water vapor pointed to a beautiful simplicity. On repeating the experiment most carefully, he found that the *combining volumes* of hydrogen and oxygen are as 2 to 1, within 0.1%. Now, Gay-Lussac remembered that gases as a group exhibit simple and regular physical laws, i.e., the so-called gas laws of Boyle and of Charles, which are the same for all gases, and he speculated that therefore similarly simple results might be obtained in the combination of other gases. From his analysis of previously published work and his own careful research he concluded, "It appears to me that gases always combine in the simplest proportions when they act on one another; and we have seen in reality in all the preceding examples that the ratio of combinations is 1 to 1, 1 to 2, or 1 to 3." (To this we should now add other simple ratios by volume, such as 2 to 3.) Furthermore, he noted that if the compound resulting from the chemical reaction remains a gas or a vapor at the temperature of the experiment, the volume of the compound is also simply related to the volumes of the participating elements. We might write results of this kind as follows:

1 vol nitrogen gas + 3 vol hydrogen gas = 2 vol ammonia vapor,  
2 vol hydrogen gas + 1 vol oxygen gas = 2 vol water vapor,  
1 vol nitrogen gas + 1 vol oxygen gas = 2 vol nitric oxide gas,  
1 vol hydrogen gas + 1 vol chlorine gas = 2 vol hydrochloric acid  
gas.

This marvelous simplicity would appeal to us at first sight, but for many reasons Dalton rejected these results utterly and to the very end. We recall that his model of gases, in order to explain the homogeneity of the atmosphere, required that the atoms of different elements be of different sizes. To accept the possibility of such general simplicity, even occasionally of *equal* combining volumes, seemed to him tantamount to accepting the notion that equal gas volumes contain equal numbers of atoms, or at any rate multiples thereof,



Amedeo Avogadro

FIG. 19-3. Amedeo Avogadro (1776-1856).

that the atoms therefore might be of equal sizes, and that consequently he would have to give up both his static model and what he supposed to be Newton's "immutable Law of elastic fluids." His arguments were aided by the fact that his own faulty volumetric experiments contradicted Gay-Lussac's results, and that these, of course, were admittedly not *exactly* whole numbers (we recall that scientific measurements can never be made without some error, and in these cases the likelihood of error was often considerable). Therefore Gay-Lussac had had to round off his figures, sometimes by a few percent, to obtain his simple results, and to Dalton this small discrepancy was not a negligible experimental error but the most significant truth. This dispute summarizes for us brilliantly and compactly some typical features of the life of science: the search for simplicity, the fascination of integers, the inexactitude of measurement, the role of a theory to guide and interpret experimental results, and so forth; indeed, we are here reminded a little of the dispute between Galileo and the scholastic regarding the interpretation of free-falling motion.

Quite apart from Dalton's objections, Gay-Lussac's results do

not yet lend themselves directly to an interpretation by Dalton's atomic theory of gases. To make the two dovetail required an amendment of the theory, after which, however, the atomic hypothesis emerged stronger than ever, in fact, approximately in its present state. And this amendment was not long in forthcoming; it was the work of the Italian physicist Amedeo Avogadro (Fig. 19-3).

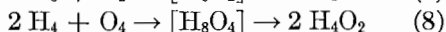
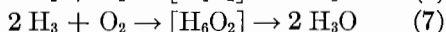
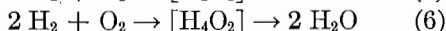
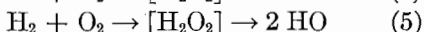
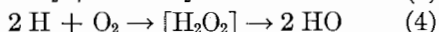
**19-6 Avogadro's hypothesis.** Avogadro's obscurely phrased but penetrating publication of 1811 may be put into four separate parts:

(a) As had been often and persistently suggested by Gay-Lussac and others, Dalton's picture of a gas in which the particles are in contact is now replaced by another (it does not matter whether a static or a kinetic model) in which the particles have only a very thin shell of caloric and are quite small compared with the distances between them. The volume actually taken up by the atoms is consequently very much smaller than in Dalton's model, in fact, a negligible fraction of the total volume of any sample. It is a return, if you wish, to Lucretius' "atom and void" picture of the gas. With this assumption we must, of course, also give up the hope of obtaining any information on *relative atomic dimensions* on the basis of combining volumes. But this problem of dimensions, so important to Dalton, immediately becomes quite insignificant.

(b) Contrary to initial considerations of simplicity, or to the firm faith of Dalton, or, in fact, to every reasonable expectation at the time, the ultimate particles of gaseous *elements* are conceived to be not necessarily just single solitary atoms, but instead may be thought to be made of two or more chemically alike atoms "united by attraction to form a single whole." Calling the smallest such unit a *molecule*, we may then imagine that the molecules of element *X* in gaseous form all consist of two atoms each (a diatomic gas), or in the case of some element *Y* of three atoms (triatomic), or in gas *Z* of single atoms (monatomic), and so on. We still bear in mind that since the molecules of an element are made up of, at most, a few atoms they are negligibly small compared with the intermolecular distances.

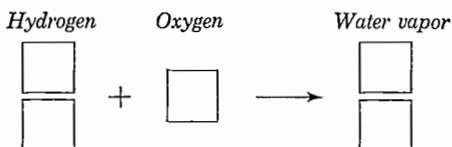
(c) In a *gaseous compound*, e.g., water vapor, the molecules evidently each consist of definite numbers of chemically unlike atoms. But when two gases combine, as in the formation of water from hydrogen and oxygen gas, we must not just *assume* that one or more molecules of hydrogen and one or more of oxygen simply associate into a new or larger unit, we must examine the possibility that such a larger group forms only for a brief moment and then at once breaks

up into two or more equal parts. We may represent the various possibilities symbolically as follows:

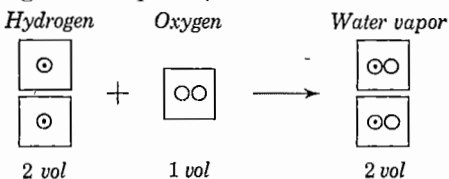


*etcetera.*

(d) But the most important part of Avogadro's proposal is that at equal temperatures and pressures (the conditions largely obeyed in Gay-Lussac's experiments) *equal volumes of pure gases, whether elements or compounds, do contain equal numbers of molecules.* To turn to specific cases, let us first redraw Gay-Lussac's results for water in symbolic form:



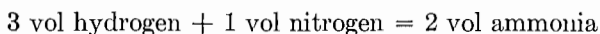
Each of these squares represents a unit volume (say 1 cm<sup>3</sup>). In accord with Avogadro's postulates, we must draw into each square the same number of molecules; for simplicity, let this be just one molecule. If with Dalton we now assume that hydrogen and oxygen gases are monatomic, and if the reaction proceeds according to equation (1) above, we should at once find it impossible, without splitting the very atom of oxygen, to place into each of the two volumes of water vapor one whole molecule of the same material. We must also reject equations (2) and (3) for a similar reason. If we really believe in Avogadro's hypothesis deeply enough to entertain the possibility that one of the other equations represents the true reaction, we may next turn to the assumption that hydrogen is indeed monatomic but that oxygen is diatomic; then equation (4) applies, and filling in our squares,



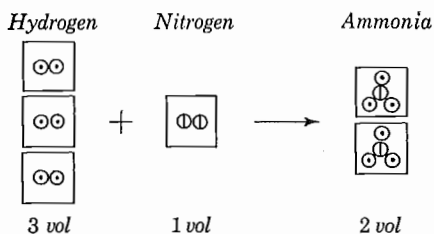


a reaction that is evidently in harmony both with the expectations of Avogadro's hypothesis and Gay-Lussac's experimental results.

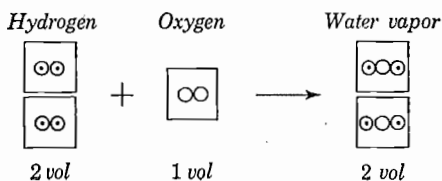
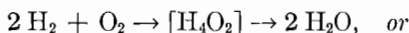
But alas! equations (6) and (7), and many others that may be imagined, would *also* be consistent with these two factors. [Demonstrate this as a problem for the cases of equations (6) and (7), and for one other reaction equation that could work.] Evidently there is an unlimited number of such possibilities; in abandoning Dalton's too simple reaction  $H + O = HO$  as wrong, we have discovered too many alternatives that may all be right. However, we can at once exclude some of these by considering how these gases behave in other reactions. From Gay-Lussac's experiments we know of the following volume-relation in the production of ammonia vapor:



This result shows us at once that hydrogen cannot be monatomic if each unit volume of ammonia is to have the same number of molecules as each unit volume of hydrogen (demonstrate). But the experimental result harmonizes with Avogadro's hypothesis *if hydrogen is diatomic* and if the same applies to nitrogen; for in this case we may write:  $3 H_2 + N_2 \rightarrow [N_2H_6] \rightarrow 2 NH_3$ , or



Having now ruled out that hydrogen gas is monatomic, we have automatically denied equation (4) for the production of water vapor. Equation (5) is impossible because it conflicts with Gay-Lussac's results on the combining volumes (demonstrate). So equation (6) emerges as the simplest one that may still be correct:



At this point Avogadro himself, quite properly, assumes also a "rule of simplicity"; that is, he regards as the correct process the one that is consistent with other chemical reactions, simplest, *and* in harmony with Gay-Lussac's results and his own postulate, thereby choosing equation (6) over (7), (8), and all the others that may be constructed along these lines. But this is—and was—perhaps a bit disappointing; for all along we had hoped to do without any arbitrary rules of this kind. However, there is this difference: Dalton's rule, by assigning molecular formulas based on fairly incomplete data, had led at once to conflicts, whereas Avogadro's rule of simplicity may be used as a final criterion of choice *after* all the known conflicts have been considered.

PROBLEM 19-3. Show that the reaction  $\text{N}_2 + \text{O}_2 \rightarrow 2 \text{NO}$  is consistent both with Gay-Lussac's volumetric experiment on nitric oxide and with Avogadro's hypothesis. Why are we sure that this reaction cannot proceed according to the equation  $\text{N} + \text{O} \rightarrow \text{NO}$ ? If the production of nitrous oxide ( $\text{N}_2\text{O}$ ) is given by the equation  $2 \text{N}_2 + \text{O}_2 \rightarrow 2 \text{N}_2\text{O}$ , what must be the volumetric relations in this reaction?

In summary, let us note the following features of Avogadro's accomplishment.

1. He, unlike Dalton, accepted Gay-Lussac's results as correct and showed that the work of both investigators was mutually consistent.

2. He replaced the static and atomic model of gases by the kinetic and molecular one.

3. He did not supply any physical reason for his radical postulate that equal volumes of gases contain equal numbers of molecules, other than that this scheme would help to explain gaseous reactions—although Dalton, on the other hand, had a good physical reason in his simple static model for rejecting Avogadro's idea.

4. In postulating molecules, Avogadro could offer no plausible explanation for the proposition that chemically similar atoms in a gas cling together in twos or threes, etc. If such forces of attraction could exist between like atoms, why, it was asked, do these forces not cause the whole bulk of the gas to condense? On such grounds Dalton had, in 1809, rejected as absurd the notion of gaseous elements other than monatomic. Nor did others take more kindly to the idea. Chemists after 1811 began to become more and more convinced that the stability of molecules such as those of water could be explained by strong positive or negative charges attached

to the atoms of different elements; therefore it was difficult to visualize why two hydrogen atoms, which evidently would present each other with equal charges, should still be firmly bound together in a molecule. As a matter of fact, it has been possible to answer such questions only within the last forty years.

5. Later on other elements could be prepared in gaseous form at higher temperatures, e.g., mercury, phosphorus, arsenic, and sulfur. On the basis of Avogadro's hypothesis, the chemical and physical behavior of these gases could be accounted for only by assuming that the molecules of these gases were *not* all diatomic, but that, instead, the molecular formulas were  $Hg$ ,  $P_4$ ,  $As_4$ , and, depending on the temperature,  $S$ ,  $S_2$ ,  $S_6$ , and  $S_8$ . At the outset there had been only four elements known in gaseous form, namely, nitrogen, oxygen, hydrogen, and chlorine, and for chemical reactions involving these, Avogadro's hypothesis worked well provided diatomic molecules for each were assumed. Some compromise might eventually have been made if all elements had turned out to be diatomic. But now it became apparent that this was not the case. Why this diversity? It seemed to many that Avogadro, while trying to achieve simplicity, had produced confusion. Actually, an independent check on the number of atoms in a molecule of gaseous material was made possible only much later by the kinetic theory of heat, to be discussed in the following chapter.

6. There seemed to be little plausibility attached to the idea—and less evidence for it—that during the reaction of gases intermediate unstable molecules were formed which at once broke into two or more equal fragments. Although we know today that the intermediate stages in such reactions are far more complex, it is still true that Avogadro's remarkable intuition is closer to the facts than the more naive conception of his contemporaries.

7. Avogadro's postulate that all pure gases at the same pressure and temperature have the same number of molecules per cubic centimeter did not allow him to discover the *exact value* of that number—other than the vague suggestion that it would be very large. Experiments in many separate branches of physical science since that time have indirectly produced values, and the evidences all agree fairly well on about  $2.69 \times 10^{19}$  molecules per  $cm^3$  of any pure gas at  $0^\circ C$  and 1 atmosphere pressure. Larger volumes would contain proportionally larger numbers.

8. At the time, two not inconsiderable obstacles to the immediate acceptance of Avogadro's work were the obscure manner of present-

ation and the large admixture of other daring—but erroneous—conclusions.

**19-7 Chemistry after Avogadro.** Some of these points may make plausible why Avogadro's work lay neglected for almost half a century, during which time chemistry, initially spurred on by Dalton's contribution, progressed vigorously. But this very activity had an inevitable result: it unearthed more and more evident data testifying to an insufficiency in Dalton's speculations. Consequently, while Avogadro's contribution remained in obscurity, other possible amendments to Dalton's conceptual schemes were strongly debated. They all, however, suffered from fundamental flaws, e.g., the persistent assumption that the atoms in a gas are contiguous. The discouragement of unavoidable contradictions seems to have produced in the 1840's a gradual loss of faith in the whole atomic theory of chemistry.

The solution to the dilemma came after 1858 when the Italian chemist Stanislao Cannizzaro called for a return to Avogadro's work and showed, partly through his own experiments, that it provided indeed a rational basis for chemistry. In fact, this was the final stage of the revolution of chemistry—thus were achieved the fundamental notions on which modern chemistry is built. But if Cannizzaro was successful, his persuasiveness had a powerful ally. By that time chemists had become more prepared to accept the ancient suggestion, the kinetic model of a gas, for in the meantime, a revolution had begun in physics—the abandonment of the caloric fluid, and the growing acceptance of the kinetic model of gases in that field.

Before we turn to an account of that work, let us survey several important positive extensions of Dalton's work resulting from Avogadro's hypothesis.

(a) Instead of Dalton's rule of greatest simplicity, we now have a new basis for the determination of chemical formulas, although, of course, directly applicable only to gaseous reactions.

**PROBLEM 19-4.** Accepting Avogadro's hypothesis, show that the following experimental result is harmonious with the proposition that hydrogen and chlorine gas are both diatomic, but is in contradiction to the early idea that hydrogen and chlorine gas are monatomic:

1 vol hydrogen gas + 1 vol chlorine gas = 2 vol hydrochloric acid vapor.

**PROBLEM 19-5.** A frequently used system for generating hydrogen on a

large scale in a portable plant involves the following reaction: a unit volume of alcohol treated with 1 vol steam produces 1 vol carbon dioxide and 3 vol hydrogen gas. Prove that consequently the molecular formula of the alcohol is  $\text{CH}_3\text{OH}$  (wood alcohol).

(b) Given a more rational procedure for obtaining molecular formulas, we may at once expect an extension and correction of the table of relative atomic weights. For gases the method is extremely simple. If, for example, we compare (at the same temperature and pressure) the *relative weights of equal volumes* of several gases, we would find by experiment the following relative masses: 2.016 for hydrogen, 28.02 for nitrogen, 32.000 for oxygen gas. If we agree to refer all relative weights to 32.000 for oxygen gas, we may call

Table 19-1. Relative Molecular Weights \*

<i>Gas or Vapor</i>	<i>Relative Molecular Weight (M)</i>	<i>Molecular Formula</i>
Elements:		
Hydrogen	2.016	$\text{H}_2$
Helium	4.002	He
Nitrogen	28.02	$\text{N}_2$
Oxygen	32.000 (Definition)	$\text{O}_2$
Fluorine	38.00	$\text{F}_2$
Argon	39.94	A
Chlorine	70.92	$\text{Cl}_2$
Phosphorus	123.92	$\text{P}_4$
Mercury	200.61	Hg
Compounds:		
Water vapor	18.02	$\text{H}_2\text{O}$
Carbon monoxide	28.01	CO
Carbon dioxide	44.01	$\text{CO}_2$
Uranium hexafluoride	352.07	$\text{UF}_6$
Mixture:		
Air	29 (approximate average)	$\text{N}_2$ (78%), $\text{O}_2$ (21%), $\text{CO}_2$ (0.03-0.04%), also water vapor, rare gases, etc.

\* In a less strict sense of the word the term molecular weight is used even for gaseous mixtures (as for air), or for the molecules of a solid or liquid where the molecules may or may not exist in groups, and where molecular formulas, consequently, hardly have the same meaning as in the case of gases.

these numbers *relative molecular weights* ( $M$ ). On this basis we may assemble a table of relative molecular weights of some gas molecules—for elements as well as compounds (Table 19-1).

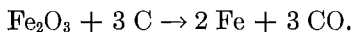
The *relative atomic weights* are, of course, obtained at once by dividing the molecular weight of the element (*cf.* Table 19-1) by the number of atoms per molecule; the required molecular formulas are obtained from the nature of chemical reactions, particularly from the volumetric relations of gases, as has been demonstrated. Clearly, the standard for relative atomic weights is the oxygen atom (16.000); a complete list of modern values for all elements will be found in the table of elements (Appendix IV).

PROBLEM 19-6. What should be the relative molecular weight of wood alcohol (methyl alcohol,  $\text{CH}_3\text{OH}$ )? Ethyl alcohol ( $\text{C}_2\text{H}_5\text{OH}$ )? Diethyl ether [ $(\text{C}_2\text{H}_5)_2\text{O}$ ]? Quinine ( $\text{C}_{20}\text{H}_{24}\text{N}_2\text{O}_2$ )? Glucose ( $\text{C}_6\text{H}_{12}\text{O}_6$ )? Clay [ $\text{Al}_2\text{O}_3 \cdot 2(\text{SiO}_2) \cdot 2(\text{H}_2\text{O})$ ]?

PROBLEM 19-7. Did our chemical evidences exclude the possibility that hydrogen is triatomic? What type of experiment would you look for in order to prove or disprove the hypothesis that hydrogen is triatomic? If that hypothesis should be found to be correct, what do the experiments cited in this chapter tell us about the molecular structure of the other gases? How would the molecular weights given above be modified?

(c) In problem 19-2 you obtained exclusively from the internal consistency of data on combining weights the probable molecular formulas of five compounds. Note that of these, two were solids, one liquid, and two gaseous. Later we saw that the work of Gay-Lussac and Avogadro supplemented and much simplified such calculations, at least for gaseous reactions. These approaches evidently made possible a complete determination of all atomic weights. Now we must make more explicit the relationship between these three important factors in any reaction, namely, the combining weights, the formula for the reaction, and the atomic and molecular weights of the substances involved. *Given any two of these, the third can be calculated.* These two examples will help to illustrate this point.

*Example 1.* In the reduction of the black iron ore, hematite, iron is obtained by burning carbon with the ore in the following reaction:



Given this reaction equation, and the atomic weights ( $\text{Fe} = 55.85$ ,  $\text{O} = 16.00$ ,  $\text{C} = 12.01$ ), find the combining weights.

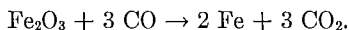
*Solution:* The following must represent the proportions by weight of the substances involved:

$$\begin{array}{rcc}
 \text{Hematite} & & \text{Carbon} \\
 \underbrace{[(55.85) \times 2 + (16.00) \times 3]} & + & \underbrace{[3 \times 12.01]} \\
 159.70 & + & 36.03 \\
 & \rightarrow & \begin{array}{r} \text{Iron} \\ \underbrace{[2 \times 55.85]} \end{array} + \begin{array}{r} \text{Carbon monoxide} \\ \underbrace{[3 \times (12.01 + 16.00)]} \end{array} \\
 & \rightarrow & 111.70 + 84.03
 \end{array}$$

This schedule informs us that if we hope to obtain about 112 gm of iron, we must supply the furnace with about 160 gm ore, 36 gm pure coal (or equivalent), and must take precautions to dispose of 84 gm of carbon monoxide gas.

PROBLEM 19-8. How many tons of carbon must we supply if we expect to obtain 80 tons of iron?

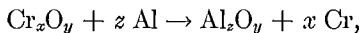
PROBLEM 19-9. Depending on the temperature of the furnace, another possible reaction is



What are the relative weights of the participants in this reaction?

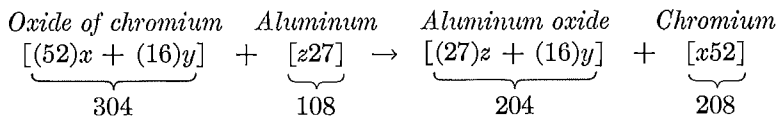
*Example 2.* When 304 gm of a certain oxide of chromium are ignited with 108 gm aluminum (Al, at. wt. about 27), there is formed an aluminum oxide and 208 gm of the pure metal chromium (Cr, at. wt. about 52). From these data, find the equation for this reaction.

*Solution:* First of all, there must be  $(304 + 108) - 208$  gm of aluminum oxide, i.e., 204 gm. Now, we do not know the molecular formulas, but they can be written for the time being as follows,



where  $x$ ,  $y$ , and  $z$  may each stand for a whole integer (1, 2, 3, . . .), for evidently the equation, if it refers to combinations of atoms, must contain as many atoms of each element on one side as on the other; that is, the equation must *balance*.

We may now regard the relative combining weights as indicative of the relative molecular weights, and say



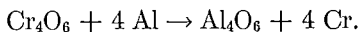
$$z \cdot \text{Al} = 108 \quad \text{or} \quad z(27) = 108; \quad z = \frac{108}{27} = 4,$$

$$x \cdot \text{Cr} = 208 \quad \text{or} \quad x(52) = 208; \quad x = \frac{208}{52} = 4.$$

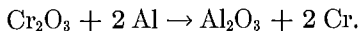
Therefore,

$$\text{Cr}_x\text{O}_y = \text{Cr}_4\text{O}_6 = [(4 \times 52) + (y \times 16)] = 304; \quad y = \frac{304 - 208}{16} = 6.$$

Our reaction equation is then



More simply,  $2 \text{Cr}_2\text{O}_3 + 4 \text{Al} \rightarrow 2 \text{Al}_2\text{O}_3 + 4 \text{Cr}$ , and dividing through by 2, finally



PROBLEM 19-10. At one time, aluminum was produced by reducing aluminum chloride with metallic sodium (Na, at. wt. about 23), yielding sodium chloride (NaCl) and aluminum (Al, at. wt. about 27). If 34.5 gm sodium are needed to obtain 13.5 gm aluminum from 66.8 gm aluminum chloride, what is the simplest reaction formula? (Note: To obtain  $x$ ,  $y$ , and  $z$  in integral form, one may have to multiply all these given quantities by some one number.)

(d) Finally, we should mention a fruitful concept that makes these calculations very much more simple: the *gram-molecular-weight* or "mole." By definition, one gram-molecular-weight, or 1 mole, of any substance is an amount whose mass is numerically equal to its relative molecular weight. Thus, 1 mole of oxygen is nothing but 32.000 gm of oxygen; 1 mole of hydrogen = 2.016 gm of hydrogen; 3 moles of mercury = 601.83 gm of mercury (i.e.,  $3 \times 200.61$ ; note that for monatomic substances the term gram-molecular-weight is often replaced by "gram-atomic-weight");  $n$  moles of water =  $n \times 18.016$  gm of water. Conversely, if a substance is known to have the molecular weight  $M$ , then  $y$  gm of that substance correspond to  $y/M$  moles.

It stands to reason that *all pure substances must have the same number of molecules in one mole*, no matter what the state, temperature, or pressure. The proof is simple: The total mass of a substance



( $m$ ) is the mass per molecule ( $m_0$ )  $\times$  number of molecules ( $N_m$ ), or  $m = m_0 N_m$ . If the mass is taken to be 1 mole, it is numerically equal to  $M$ , the molecular weight; then  $M = m_0 N_m$ . If we compare two substances with molecular weights  $M'$  and  $M''$ , we may write

$$\left. \begin{aligned} M' &= m'_0 N'_m \\ M'' &= m''_0 N''_m \end{aligned} \right\}$$

Taking the ratio,  $\frac{M'}{M''} = \frac{m'_0}{m''_0} \cdot \frac{N'_m}{N''_m}$ . But by definition the ratio  $m'_0/m''_0$  is the relative molecular weight of one substance with respect to the other, or  $M'/M'' = m'_0/m''_0$ . Thus  $N'_m = N''_m$ , as was to be proved. Furthermore, the actual number of molecules per mole, by many indirect tests, is  $6.02 \times 10^{23}$  (called *Avogadro's number*, symbolized by  $N$ ).

We recall that at  $0^\circ\text{C}$  and 1 atm pressure, all pure gases contain  $2.69 \times 10^{19}$  molecules per  $\text{cm}^3$ . It is now evident that under these standard conditions,  $6.02 \times 10^{23}$  molecules would occupy a volume of  $(6.02 \times 10^{23}) / (2.69 \times 10^{19}) = 22.4 \times 10^3 \text{ cm}^3 = 22.4$  liters. This brings us to a common form of stating Avogadro's hypothesis: *At  $0^\circ\text{C}$  and 1 atm pressure, one mole of any pure gas, containing  $6.02 \times 10^{23}$  molecules, takes a volume of 22.4 liters.*

A useful and immediate application of the concept "mole" springs from this argument: If we know that in a certain reaction every molecule of  $A$  joins with  $x$  molecules of  $B$ , then we must provide for each mole of  $A$  exactly  $x$  moles of  $B$  (or corresponding fractions) to assure a complete reaction.

*Example 3.* Of a substance  $X$  we are told that it is formed when equal numbers of moles of two elements combine. When a quantity of 40 gm of  $X$  is analyzed, it is found to contain about 35.5 gm chlorine and 4.5 gm of an unidentified element  $Y$ ; the stated amount of chlorine is  $\frac{1}{2}$  mole of the chlorine gas  $\text{Cl}_2$ , because

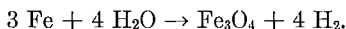
$$n = \frac{m}{M} = \frac{35.5 \text{ gm}}{71 \text{ gm/mole}} = \frac{1}{2};$$

therefore the 4.5 gm of  $Y$  represent  $\frac{1}{2}$  mole of  $Y$  also. This makes the molecular weight of  $Y \doteq 9$ . If  $Y$  is monatomic, this is also its atomic weight; if diatomic, the atomic weight is 4.5; if triatomic, 3; etc. A glance at the table of atomic weights shows that, in fact,

only one of these choices exists—namely, beryllium (Be) with an atomic weight 9 (or more accurately, 9.02). Thus we have identified the substance  $X$  as  $\text{BeCl}_2$ .

**PROBLEM 19-11.** How many moles of ammonia is 1 gm of that vapor? How many molecules are there in this quantity? How many *atoms*? What would be its volume at  $0^\circ\text{C}$  and 1 atm? at  $100^\circ\text{C}$  and 5 atmospheres? (Use the general gas law.) How many moles of hydrogen does this vapor contain?

**PROBLEM 19-12.** The reaction representing the rusting of iron may be written



How many moles of hydrogen are liberated for each gram of iron completely rusted?

**PROBLEM 19-13.** Knowing Avogadro's number, the molecular weights of certain gases, and their molecular formulas, calculate the "absolute" mass (in grams) of a single atom of hydrogen, of oxygen, and of chlorine.

**PROBLEM 19-14.** Evaluate Dalton's atomic theory of chemistry from the point of view of our criteria for theories in Chapter 8.

**PROBLEM 19-15.** Discuss the growth of the atomistic hypothesis to this point in the light of Chapter 12.

**19-8 Valence.** Dalton's atomic theory suggests a very simple interpretation of the empirical law of definite proportions. Carbon monoxide, for example, always consists of molecules made of one carbon atom and one oxygen atom each; each molecule of ammonia invariably consists of one nitrogen atom in close connection with three hydrogen atoms; and similarly for each molecule of any other compound. However, contemplation of this gratifying regularity soon brings up perplexing questions: Why do atoms have such fixed combining powers with one another? What determines the fact that hydrogen and chlorine always make  $\text{HCl}$  and not  $\text{HCl}_3$  or  $\text{H}_5\text{Cl}_2$ ? Why, given H and O, can we find either  $\text{H}_2\text{O}$  or  $\text{H}_2\text{O}_2$ , but not  $\text{H}_2\text{O}_3$ , or, given C and O, we find  $\text{CO}$  or  $\text{CO}_2$  but not  $\text{C}_2\text{O}$ ?

Familiarity with the composition of many molecules might lead to a picturization of hydrogen atoms as having one "hand" or "hook" each by which to engage a similar "hook" on other atoms, as indicated by the formulas  $\text{HCl}$  and  $\text{H}_2\text{O}$ . Thus one hook should be assigned to the chlorine atom, two to oxygen atoms (see Fig. 19-4). The difference between the structure of  $\text{H}_2\text{O}$  and that of  $\text{H}_2\text{O}_2$  is then directly visualizable. The formula  $\text{CO}_2$  implies that the carbon atom has four such coupling mechanisms to engage the two of each

oxygen atom; and in CO it would seem that two hooks of carbon are left unoccupied and ready to snatch more oxygen—a simple picture that coincides with the fact that CO is such a powerful reducing agent.

Ultimately this type of analogy is found to be incomplete and dangerous. The question of combining powers had to be answered in terms of the structure of the atoms themselves. In the meantime, this general approach formalized the subject in a powerful way, even though it remained at first mainly a descriptive device. Thus we encounter, as of 1852, the concept of *valence*—now indispensable for the description and understanding of chemical processes. The valence of an element, speaking very roughly, is the number of “hooks” which an element can put out to enter into chemical partnership with other elements or with stable groups of atoms (radicals). The simplest way to determine the valence of an element is to look up the chemical formula of a substance involving only this element and hydrogen, which we regard *a priori* as having a valence of 1 (monovalent). For example, the formulas  $H_2O$ ,  $NH_3$ ,  $CH_4$  indicate valence numbers of  $O = 2$ ,  $N = 3$ ,  $C = 4$ . The formula  $H_2SO_4$  shows that the sulfate radical ( $SO_4$ ) has the valence of 2 (divalent). Next we may find further valences by noting the number of hydrogen atoms an element can *displace* in a molecule, e.g., in  $CaO$ , Ca takes the place of 2 hydrogen atoms in  $H_2O$ , hence Ca has the valence 2. The occurrence of  $HF$  (hydrofluoric acid) and of  $OsF_8$  (osmium octafluoride) shows that osmium can have a valence of 8, the highest valence number that exists. Similarly, from  $NaCl$ ,  $CaCl_2$ ,  $AlCl_3$ ,  $SnCl_4$ ,  $PCl_5$ , and  $WCl_6$  we derive valences for  $Na = 1$ ,  $Ca = 2$ ,  $Al = 3$ ,  $Sn = 4$ ,  $P = 5$ , and  $W = 6$ .

This process of enumeration, however, leads to difficulties. Carbon is quadrivalent and oxygen is divalent by the above test. Yet we just saw that they do form CO. There the total number of available “hooks” is not engaged, and the valences of the parts of the molecule do not balance. We must now say either that two of the four valence bonds of C are not used in CO, or that the valence of

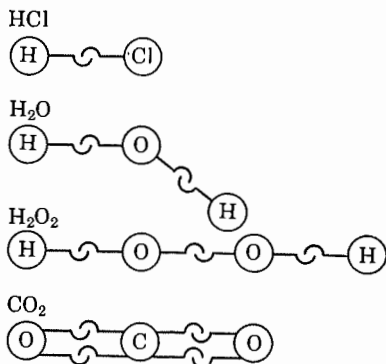


FIG. 19-4. Schematic representation of bonding in molecules.

carbon is effectively 2 in CO but 4 in CO<sub>2</sub>. In the long run the latter mode of expression is preferable, although it leads to the disappointing realization that carbon and many other elements can take on different valences depending on what molecule they help to form. The oxides of nitrogen (N<sub>2</sub>O, NO, N<sub>2</sub>O<sub>3</sub>, NO<sub>2</sub>, N<sub>2</sub>O<sub>5</sub>) indicate that nitrogen can take on the valence numbers 1, 2, 3, 4, and 5, respectively (O = 2 throughout).

By the same argument, the valence of iron atoms is 2 in FeO, FeCl<sub>2</sub>, and Fe(NO<sub>3</sub>)<sub>2</sub>, but is 3 in Fe<sub>2</sub>O<sub>3</sub>, FeCl<sub>3</sub>, and Fe(NO<sub>3</sub>)<sub>3</sub>. Uranium can have five different valences, chlorine six, rhenium seven. Historically, this element of variability brought on a great deal of debate throughout the last half of the 19th century.\*

Of the many valuable consequences of the valence concept, one is particularly noteworthy at this point. The device of valence bonds, in essence the same as the hooks used in Fig. 19-4, supplemented Dalton's pictorial representation of the atoms themselves and invited speculation on the structural arrangement in the molecule, particularly for the complex organic substances where the mind needed such a crutch badly. This was all the more urgent because there had been discovered sets of *isomers*,\*\* i.e., substances with precisely the same chemical formula but with very different physical and chemical properties, and the only hope for explaining the differences was the hypothesis of different placement of the atoms within the molecule. Hence the rise of the graphic structural formulas for representing chemical compounds such as those in Fig. 19-5 for C<sub>2</sub>H<sub>6</sub> (ethane), C<sub>2</sub>H<sub>6</sub>O (ethyl alcohol, better written C<sub>2</sub>H<sub>5</sub>OH), the famous case of C<sub>6</sub>H<sub>6</sub> (benzene; note the closed ring), and the isomers C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>, one being glucose (also called dextrose and grape sugar) and the other fructose (or levulose, fruit sugar).

But in the 1860's, when such representation first became widely used among chemists, there was hesitation in regarding them initially as anything more than working schemes without necessary relation to the actual three-dimensional and essentially inaccessible structure of molecules. In truth, although the structural formula is indispensable as a rough sketch of nature's plan for the molecule, the correct disposition of atoms and stable groups of atoms in the molecule of a compound is generally difficult to determine even with

---

\* The concept of valence brought other troubles also. For instance, a compound sometimes has an anomalous valence number, as in Fe<sub>3</sub>O<sub>4</sub>. It is probable that this is a mixture of two valence states, FeO and Fe<sub>2</sub>O<sub>3</sub>.

\*\* From the Greek *isos* + *meros*, composed of equal parts.

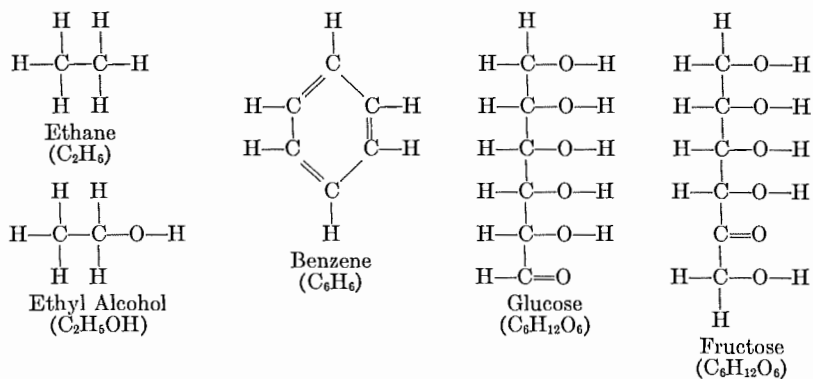


Fig. 19-5. Structural formulas for some organic compounds.

our present powerful tools of research. And even if the structure of all compounds were known with greatest accuracy, we would still have no clue to the initial puzzle of why atoms combine in such limited and specific compounds and what the physical mechanism is that holds the components of a compound together. These questions, like so many related ones, were answered only much later after the physical structure of the atom became clear (and it is during the discussion of that problem in Part G that we shall return to this topic). In the meantime, the concept of valence played a dominant role in the shaping of another conceptual scheme of profound importance—the classification of the elements.

**19-9 The search for regularity among the list of elements.** As the early list of chemical elements became longer and as their physical and chemical properties became increasingly better established, there arose rather naturally the urge to find some relation among the elements that would bring them into a logical sequence—a longing analogous to the Keplerian quest for principles that would relate the separate planets.

At first glance, the task would seem hopeless. The most immediately striking thing is that each element seems to have unique properties that differentiate it from each of the other elements. Most are solids under ordinary circumstances, but many are gases and some are liquids; their relative abundances on earth have widely different values, and so have their physical characteristics—boiling point, density, and the like. Nevertheless, there are some clues for speculation. In 1815-1816 the English physician William Prout had suggested that the atoms of all elements were but the condensation

of different amounts of hydrogen atoms, which themselves were the true fundamental particles. Consequently the atomic weights could be expected to be whole multiples of the atomic weight of hydrogen; and indeed the early determination of atomic weights by Dalton did allow this hypothesis. However, more accurate measurements were soon forthcoming, particularly those of the great chemists Berzelius, Dumas, and Stas, all demonstrating real deviations from Prout's rule. A glance at a modern and accurate table of elements shows that only a few of the atomic weights are even close to being exact multiples of the atomic weight of hydrogen. Thus Prout's hypothesis, despite its direct appeal, had to yield to the contrary evidences of quantitative analysis.

But at the same time this type of research favored another trend of thinking to bring order into the apparently haphazard list of elements, namely, the attempts to connect the atomic weights of elements with their chemical properties. It had long been known that dispersed throughout the list of elements are *families* of elements of similar chemical behavior. An example of such a family is that of the *halogens*—a group that includes the elements fluorine (F), chlorine (Cl), bromine (Br), and iodine (I); their respective atomic weights are approximately 19, 35.5, 80, 127. Although they exhibit some striking dissimilarities (e.g., the first two are gases, the third ordinarily a liquid, the last a volatile solid), these four elements have much in common. They all combine with many metals to form white, crystalline salts (*halogen* means "salt-former"). The boiling point changes progressively as we shift attention from the first to the last one (F,  $-187^{\circ}\text{C}$ ; Cl,  $-35^{\circ}\text{C}$ ; Br,  $59^{\circ}\text{C}$ ; I,  $184^{\circ}\text{C}$ ), and similarly for their melting points. The solubility in water progressively decreases, and so does the chemical activity, as manifested by the violence of chemical reactions (speed of reaction, amount of energy liberated).

This list of properties relating the halogens by continuous changes is to be supplemented by a second list of properties which are the same for all these elements. Halogens form compounds of similar formulas in violent reaction with metals (e.g., NaF, NaCl, NaBr, NaI; or  $\text{AlF}_3$ ,  $\text{AlCl}_3$ ,  $\text{AlBr}_3$ ,  $\text{AlI}_3$ ); that is, they have the same valence in analogous reactions. With hydrogen, all four elements form simple compounds that dissolve in water and form acids. All four form a diatomic vapor under ordinary conditions.

Another family, referred to as the *alkali metals*, is formed by lithium (Li, approximate atomic weight 7), sodium (Na, 23), potas-

sium (K, 39), rubidium (Rb, 85.5), and cesium (Cs, 133). Here too, there is a long list of chemical and physical properties which are common to all five elements or which change in a constant direction with increasing atomic weight.

In the restless search for some quantitative connection between the elements, there appeared as early as 1829 the discovery of the German chemist Johann Wolfgang Döbereiner, who found that within several chemical families the atomic weight of one element had a value equal, or nearly so, to the average of its two immediate neighbors; thus 23 for Na is the average of 7 for Li and 39 for K, and in the triad of Cl, Br, and I, 80 for Br is roughly the mean of 35.5 (Cl) and 127 (I). Of course, no physical meaning could be attached to such numerical regularities (we are at once reminded of Bode's law). In fact, the underlying explanation of such relations turned out to be extremely complex (involving the stability of nuclei) and has not yet been fully given. However, the search for numerical regularities of this kind proved popular, as always, and was carried on with vigor. It is another example where research that proceeded for a time on the wrong track was nevertheless directly fruitful, for through this effort it became more evident that the list of elements did indeed have a meaningful general structure, although not according to such atomic weight groups.

From 1860 on, about the time when Cannizzaro's agitation on behalf of Avogadro's hypothesis began to succeed, it was possible in principle to fix with very good accuracy the atomic weights of almost all elements and the molecular weights of compounds. Now one could place the elements into a sequence based entirely on a property of the single atom itself, namely, its atomic weight. Such a sequence reveals a curious fact: one may discover a regular spacing between the elements belonging to one family, a recurrence of numerical position quite apart from Döbereiner's atomic weight triads. In the words of the English chemist Newlands, one of several to seek such a regularity of spacing, "the eighth element, starting from a given one, is a kind of repetition of the first, like the eighth note in an octave of music."

Newlands' attempt was, however, doomed to failure. Although it proved to be correct that chemically similar elements repeat periodically, and although it was possible to perceive several "octaves" in Newlands' arrangement, his analogies more often than not had to be rather far-fetched, for he did not foresee that the periodicity in the end would not be regular throughout the list of elements,

and he had not considered the necessity of leaving spaces for still undiscovered elements. Here you might object that prior to the discovery of the new elements one could not possibly leave gaps for them in the proper places. However, it was exactly this achievement which provided the center of the work of Mendeléeff.

**19-10 The early periodic table of elements.** In the publications in 1868 *et seq.* of the great Russian chemist Dmitri Ivanovitch Mendeléeff (1834-1907) and, starting a little later, of the German Julius Lothar Meyer (1830-1895), we find the culmination of the six decades of quantitative investigations and speculative inductions since Dalton's atomic theory first became generally known. A few of the most striking aspects of Mendeléeff's contribution can be given in his own words, even though a great deal of descriptive material between excerpts is omitted. Thus his initial paper of 1869 gives us the following background material of his work:

"During the course of the development of our science, the systematic arrangement of elements has undergone manifold and repeated changes. The most frequent classification of elements into metals and non-metals is based upon physical differences as they are observed for many simple bodies, as well as upon differences in character of the oxides and of the compounds corresponding to them. However, what at first acquaintance with the subject-matter appeared to be completely clear, lost its importance entirely when more detailed knowledge was obtained. Ever since it became known that an element such as phosphorus could appear in non-metallic as well as in metallic form, it became impossible to found a classification on physical differences. . . . In recent times the majority of chemists is inclined to achieve a correct ordering of the elements on the basis of their valency. There are many uncertainties involved in such an attempt . . .

"Thus, there does not exist yet a single universal principle which can withstand criticism, that could serve as guide to judge the relative properties of elements and that would permit their arrangement in a more or less strict system. Only with respect to some groups of elements there are no doubts that they form a whole and represent a natural series of similar manifestations of matter. Such groups are: the halogens, the alkaline earth metals, the nitrogen group, partially also the sulfur group, the companions of platinum, the companions of cerium, and a few others . . . The discovery of rubidium, cesium and thallium gave rise to the reflection that our knowledge of the elements is very limited:





A handwritten signature in cursive script, reading "D. Mendeléeff". The signature is written in dark ink on a light background.

FIG. 19-6. Dmitri Ivanovitch Mendeléeff (1834-1907).

and the attempt to construct a system appears to be premature as long as there exists no hypothetical basis which could serve as the foundation of a strict system.

“The investigations regarding the simple relations of atomic weights have caused many, in particular Dumas, Pettenkofer, Sokolow and others, to point out the numerical relations between the atomic weights of those elements which form a group; but, so far as I know, they have not led to a systematic arrangement of all known elements . . . When I undertook to write a handbook of chemistry entitled *Foundations of Chemistry*, I had to make a decision in favor of some system of elements in order not to be guided in their classification by accidental, or instinctive, reasons but by some exact, definite principle. In what has been said above we have seen the nearly complete absence of numerical relations in the construction of systems of elements; every system, however, that is based upon exactly observed numbers is to be preferred of course to other systems not based upon numbers because then only little margin is left to arbitrariness. The numerical data available regarding elements are limited at this time. Even if the physical properties of some of them have been

determined accurately, this is true only of a very small number of elements. Properties such as the optical, and even the electrical or magnetic ones, cannot serve as a basis for the system naturally, since one and the same body, according to the state in which it happens to be at the moment, may show enormous differences in this regard. With respect to this fact, it is sufficient to remember graphite and diamond, ordinary and red phosphorus . . . However, everybody does understand that in all changes of properties of elements, *something* remains unchanged, and that when elements go into compounds this material something—represents the [common] characteristics of compounds the given element can form. In this regard only a numerical value is known, and this is the atomic weight appropriate to the element. The magnitude of the atomic weight, according to the actual, essential nature of the concept, is a quantity which does not refer to the momentary state of an element but belongs to a material part of it, a part which it has in common with the free element and with all its compounds . . . For this reason I have endeavored to found the system upon the quantity of the atomic weight.

“The first attempt I undertook in this direction was the following: I selected the bodies with the smallest atomic weight and ordered them according to the magnitude of their atomic weights. Thereby it appeared that there exists a periodicity of properties and that even according to valence, one element follows the other in the order of an arithmetical sequence.

Li = 7	Be = 9.4	B = 11	C = 12	N = 14	O = 16	F = 19
Na = 23	Mg = 24	Al = 27.4	Si = 28	P = 31	S = 32	Cl = 35.3
K = 39	Ca = 40	. . . . .	Ti = 50	V = 51”	<i>(etcetera)</i>	

Note that in this arrangement hydrogen, the first and lightest of the elements, is left out, for it has rather unique properties; that helium, together with the other elements of the family of *rare gases*, had not yet been discovered; and that Mendeléeff’s values for atomic weights sometimes differ significantly from the more recent values. He writes the first seven elements from lithium to fluorine in sequence of increasing atomic weight, then writes the next seven, from sodium to chlorine, below the first line. The periodicity of chemical behavior is already obvious. In the first vertical column are the first two alkali metals, in the seventh column the first two halogens, and each of the columns between contains two chemically similar elements. Furthermore, the main valences for these columns are 1 for the first, 2 for the second, 3 for the next, then 4, 3, 2, and 1. Reading from left to right, any element (let us use a general symbol R.) in each

column may form higher oxides or hydrides according to the following scheme.

Column	1st	2nd	3rd	4th	5th	6th	7th
Higher oxide } or hydride }	$R_2O$	$R_2O_2 (= RO)$	$R_2O_3$	$R_2O_4 (= RO_2)$			
				$H_4R$	$H_3R$	$H_2R$	$HR$

When Mendeléeff now adds to the two rows of elements the third, K comes to stand directly below Li and Na, and of course we know that K is indeed a member of the same family with the same valence of unity. Next comes Ca (calcium), divalent like Be and Mg above it. Next should come the next heaviest element then known, which was titanium (Ti), and previous workers trying to develop such tables had generally done this. In Newlands' table, which almost entirely coincides with Mendeléeff's to this point, another element was placed in its stead whose atomic weight actually was much too high. This point offers us accurate insight into the dominant thoughts of Mendeléeff. He recognizes that the chemical properties of Ti require it to appear under C and Si in the fourth vertical column, not in the third column. If the classification is to be complete and meaningful, there ought to exist a hitherto unsuspected element with an atomic weight somewhere between that of Ca (40) and Ti (50, new value 47.9), and with a valence of 3. This amounts to a definite prediction, and later in the same work Mendeléeff finds other cases of this sort among the remaining elements.

The whole scheme to this point may be understood in terms of a very crude analogy. It is as if a librarian were to put all his books in one heap, weigh them individually, and then place them on a set of shelves according to increasing weight—and find that on each shelf the first book is on Art, the second on Philosophy, the third on Science, the fourth on Economics, and so on. Our librarian may not understand in the least what the underlying explanation for this astonishing regularity is, but if he now discovers on one of these shelves a sequence Art-Science-Economics, he will perhaps be very tempted indeed to leave a gap between the two books on Art and Science and look about for a missing Philosophy book of the correct weight to fill the gap.

Mendeléeff had no illusions that he understood the reason for this symmetry of arrangement, but he saw clearly that his work would eventually lead to a physical explanation, and that in the meantime

Table 19-2. Periodic classification of the elements (Mendeléeff, 1872)

GROUP→	I	II	III	IV	V	VI	VII	VIII
Higher Oxides and Hydrides	R <sub>2</sub> O —	RO —	R <sub>2</sub> O <sub>3</sub> —	RO <sub>2</sub> H <sub>4</sub> R	R <sub>2</sub> O <sub>5</sub> H <sub>3</sub> R	RO <sub>3</sub> H <sub>2</sub> R	R <sub>2</sub> O <sub>7</sub> HR	RO <sub>4</sub> —
1	H(1)							
2	Li(7)	Be(9.4)	B(11)	C(12)	N(14)	O(16)	F(19)	
3	Na(23)	Mg(24)	Al(27.3)	Si(28)	P(31)	S(32)	Cl(35.5)	
4	K(39)	Ca(40)	—(44)	Ti(48)	V(51)	Cr(52)	Mn(55)	Fe(56), Co(59), Ni(59), Cu(63)
5	[Cu(63)]	Zn(65)	—(68)	—(72)	As(75)	Se(78)	Br(80)	
6	Rb(85)	Sr(87)	?Yt(88)	Zr(90)	Nb(94)	Mo(96)	—(100)	Ru(104), Rh(104), Pd(106), Ag(108)
7	[Ag(108)]	Cd(112)	In(113)	Sn(118)	Sb(122)	Te(125)	I(127)	
8	Cs(133)	Ba(137)	?Di(138)	?Ce(140)	—	—	—	
9	—	—	—	—	—	—	—	
10	—	—	?Er(178)	?La(180)	Ta(182)	W(184)	—	Os(195), Ir(197), Pt(198), Au(199)
11	[Au(199)]	Hg(200)	Tl(204)	Pb(207)	Bi(208)	—	—	
12	—	—	—	Th(231)	—	U(240)	—	
SERIES								

“new interest will be awakened for the determination of atomic weights, for the discovery of new elements, and for finding new analogies among the elements.” At a later time he added, “just as without knowing the cause of gravitation it is possible to make use of the law of gravitation, so for the aims of chemistry it is possible to take advantage of the laws discovered by chemistry without being able to explain their causes.” The initial paper continues:

“I shall now present one of the many systems of elements which are based upon the atomic weight. They form but one attempt to represent the results which can be achieved in this direction. I am quite conscious of the fact that this attempt is not final, but it appears to me to express quite clearly already the applicability of my proposed principle to all elements whose atomic weight is determined with some reliability. Above all I was interested to find out a general system of elements.”

Table 19-2 gives Mendeléeff's arrangement in the version published in 1872. We note that Mendeléeff distributes the 63 elements then known (with 5 in doubt) in 12 horizontal lines or *series*, starting with the lightest element, hydrogen, at the top left, and ending with uranium at the bottom right. The others are written in order of increasing atomic weight, but are so placed that each element is in the nearest vertical column or *group* of chemically similar elements. Thus the halogens are in group VII, group VIII contains only ductile metals, groups I and II contain light and low melting-point metals, and the specific family of alkali metals is in group I. The more detailed arguments for each particular placement, as Mendeléeff's extensive discussion shows, are often exceedingly delicate, guided by his profound and often almost intuitive knowledge of chemical nature.

The arrangement does, however, leave many gaps: “Vacant places occur for elements which, perhaps, shall be discovered in the course of time.” And several of the elements toward the end of this table are in doubt: “The higher atomic weights belong to elements which are rarely encountered in nature, which do not form large deposits, and which therefore have been studied relatively little . . . With respect to the position of some elements that are rarely encountered in nature, there exists, quite understandably, complete uncertainty.”

But the main part of the table is fairly complete, and another feature of this scheme reveals itself: “For a true comprehension of

the matter it is very important to see that all aspects of the distribution of the elements according to their atomic weights essentially express one and the same fundamental dependence—*periodic properties.*” By this is meant that in addition to the gradual change in physical and chemical properties within each *vertical* group or family, there is also a periodic change of such properties among the elements if they are scanned in a horizontal sequence. “The properties of the elements as well as the forms and properties of their compounds are in periodic dependence on, or (to express ourselves algebraically) form a periodic function of, the atomic weights of the elements.” This *periodic law* is the heart of the matter, a generalization of Newlands’ suggestion (with whose work Mendeléeff was initially not acquainted). We can best illustrate it as Lothar Meyer did, by drawing a curve of the value taken on by some physical quantity as a function of atomic weight, for example, the “atomic volume” of each element (the ratio of atomic weight to density, the mass per unit volume, in the liquid or solid phase; Fig. 19-7). Note the irregular spacing of the “waves” yet the simultaneous periodicity of the position of analogous elements on the curve, entered here specifically for the halogens and for the alkali metals.

PROBLEM 19-16. Use the data given in the *Handbook of Chemistry and Physics* or a similar source, and plot the melting points of the elements against the atomic weights for the elements from hydrogen to barium (Ba) inclusive. Then discuss the periodicities that appear on the graph, e.g., the positions of the two families of elements mentioned thus far.

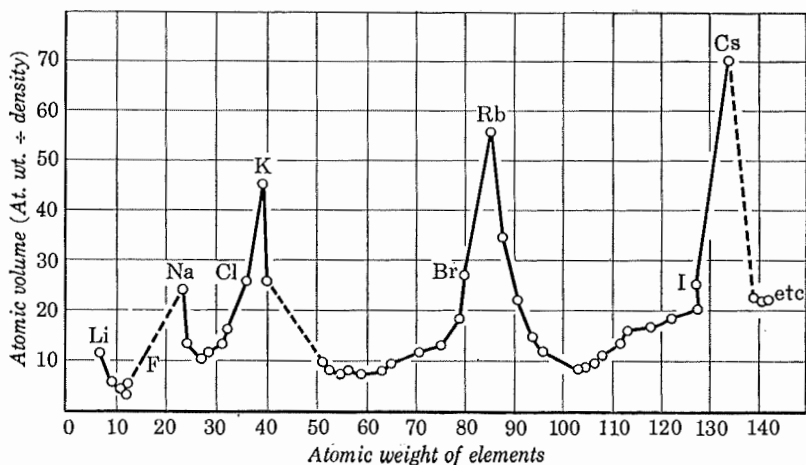


FIG. 19-7. Atomic volumes for the first portion of the Table of Elements. (After Lothar Meyer, but using modern values.)

**19-11 Consequences of the periodic law.** More detailed examination of the periodic scheme yields some additional information.

(a) If an element R enters into chemical combinations with oxygen, the composition of one of its higher oxygen compounds is determined by the *group* to which R belongs, i.e., the first group gives  $R_2O$ , the second  $R_2O_2$  (or  $RO$ ), the next  $R_2O_3$ , and so forth; similarly for combinations with hydrogen, as indicated along the top of Table 19-2.

(b) A distinction is to be made between the even and the odd series, indicated in Table 19-2 by a slight shift in the printing of the symbols to the right or left. Chemical and physical similarities are generally closer among all the elements in one group that belong to the even series only or the odd series only. Thus in group VI, S, Se, and Te are nonmetals, whereas Cr, Mo, and W are heavy metals.

(c) Each element is a member of a vertical (family) sequence and a horizontal (periodic) sequence. Thus any one property of an element is related to the rest of the table somewhat like the piece in the middle of a jigsaw puzzle or the letter in a crossword puzzle—and like it, can be induced from what is known about the neighboring regions. Here, to continue with arguments taken from Mendeléeff's textbook, is one example:

“If in a certain group there occur elements  $R_1$ ,  $R_2$ ,  $R_3$ , and if in that series which contains one of these elements, for instance  $R_2$ , an element  $Q_2$  precedes it and an element  $T_2$  succeeds it, then the properties of  $R_2$  are determined by the mean of the properties of  $R_1$ ,  $R_3$ ,  $Q_2$ , and  $T_2$ . Thus, for instance, the atomic weight of  $R_2 = \frac{1}{4}(R_1 + R_3 + Q_2 + T_2)$ . For example, selenium occurs in the same group as sulfur,  $S = 32$ , and tellurium,  $Te = 127$ , and, in the 5th series,  $As = 75$  stands before it and  $Br = 80$  after it. Hence the atomic weight of selenium should be  $\frac{1}{4}(32 + 127 + 75 + 80) = 78.5$ , which is near to that generally accepted,  $Se = 79$ , in which there is a possible error in the first decimal, so that 78.5 may be nearer the actual figure.”

We see at once to what this leads:

“. . . the periodic dependence of the properties on the atomic weights of the elements gives a *new means for determining*, by means of the equivalent weight, *the atomic weight* or atomicity of imperfectly investigated but known elements, for which no other means could as yet be applied for determining the true atomic weight. At the time [1869] when the periodic law was first pro-

posed there were several such elements. It thus became possible to learn their true atomic weights, and these were verified by later researches. Among the elements thus concerned were indium, uranium, cerium, yttrium, and others . . . Thus a true law of nature anticipates facts, foretells magnitudes, gives a hold on nature, and leads to improvements in the methods of research, etc."

(d) But there is another, much more important consequence:

"[The] properties of selenium may also be determined in this manner; for example, arsenic forms  $H_3As$ , bromine gives  $HBr$ , and it is evident that selenium, which stands between them, should form  $H_2Se$ , with properties intermediate between those of  $H_3As$  and  $HBr$ . Even the physical properties of selenium and its compounds, not to speak of their composition, being determined by the group in which it occurs, may be foreseen with a close approach to reality from the properties of sulfur, tellurium, arsenic, and bromine. *In this manner it is possible to foretell the properties of elements still unknown*, especially when it is surrounded by well-known elements. For instance, in the position IV, 5 (that is, in the IVth group and 5th series) an element is still wanting. These unknown elements may be named after the preceding known element of the same group by adding to the first syllable the prefix *eka-*, which means *one* in Sanskrit. The element IV, 5, follows after IV, 3, and this latter position being occupied by silicon, I named this formerly unknown element *ekasilicon* and its symbol  $Es$ . The following are the properties which this element should have on the basis of the known properties of silicon, tin, zinc, and arsenic. Its atomic weight is nearly 72, it forms a higher oxide  $EsO_2$ , a lower oxide  $EsO$ , compounds of the general form  $EsX_4$ , and chemically unstable lower compounds of the form  $EsX_2$ .  $Es$  gives volatile organo-metallic compounds; for instance,  $Es(CH_3)_4$ ,  $Es(CH_3)_3Cl$ , and  $Es(C_2H_5)_4$ , which boil at about  $160^\circ$ , etc.; also a volatile and liquid chloride,  $EsCl_4$ , boiling at about  $90^\circ$  and of density about 1.9.  $EsO_2$  will be the anhydride of a feeble colloidal acid, metallic  $Es$  will be rather easily obtainable from the oxides and from  $K_2EsF_6$  by reduction,  $EsS_2$  will resemble  $SnS_2$  and  $SiS_2$ , and will probably be soluble in ammonium sulphide; the density of  $Es$  will be about 5.5, and  $EsO_2$  will have a density of about 4.7, etc. . . ."

This summary of his own prediction in 1871 shows fully the genius and daring of the man, the power of his scheme. And in 1887,



C. Winckler of Freiberg did indeed discover a metal, now called germanium (Ge), having an atomic weight of about  $72\frac{1}{2}$ , a density of  $5\frac{1}{2}$ , an oxide  $\text{GeO}_2$  of density 4.7, a derivative  $\text{Ge}(\text{C}_2\text{H}_5)_4$  that boiled at  $160^\circ\text{C}$ , a liquid chloride  $\text{GeCl}_4$  with a boiling point at  $83^\circ\text{C}$  and density 1.9, and so forth. In a similar fashion Mendeléeff had described the probable specifications of the elements at group III, series 4 and 5, later found to have the predicted properties (gallium and scandium).

Even though not every aspect of Mendeléeff's work offered such marvelous agreement, these are indeed stunning results (again reminiscent of the discovery of the asteroids in the general orbits predicted for them by Bode's law). Numerical predictions of this sort are the very essence of science, and the chemical profession soon cast off its initial reluctance to embrace the main concepts of a periodic classification of elements.

**19-12 The modern periodic table.** It remains to tell briefly of some rather important rearrangements that had to be made subsequently in the periodic table. Table 19-3 is one of several modern forms, and comparison with Mendeléeff's table reveals these specific differences:

(a) Hydrogen is placed uniquely so as not to be directly associated with any single family; it does in fact have a chemical relationship with both the halogens and the alkali metals.

(b) The placement of the symbol in each rectangle is arranged so that elements along one vertical line belong to one chemical family. The elements have been given consecutive numbers, the so-called *atomic numbers*, to indicate their positions in the periodic table. Several elements have been assigned more correct atomic weights and new places.

(c) Many new elements have been entered, and there are now no more additions expected within the table. In the 57th place belongs not one element but a whole group of 14 chemically almost indistinguishable elements, the *rare earths*, written out for convenience in a separate row and given atomic numbers from 57 to 71 inclusive. These elements were largely unknown in Mendeléeff's time, and indeed they somewhat disturb the symmetry of the table. Similarly, the elements of the actinide series may best be placed into the one position marked 89.

(d) The lower end of the table has been pulled together. The horizontal rows fall into 7 periods, the first three short, the following three long ones (2 series each).

Table 19-3. Periodic table of elements

Atomic weights are the most recent adopted by the International Union of Chemistry; none are given for artificially produced elements. The full names of the elements are given in Appendix III.

Period	GROUP →		I	II	III	IV	V	VI	VII	VIII	O	
	Series											
1	1		1 H 1.0080								2 He 4.003	
2	2		3 Li 6.940	4 Be 9.013	5 B 10.82	6 C 12.010	7 N 14.008	8 O 16.000	9 F 19.00		10 Ne 20.183	
3	3		11 Na 22.997	12 Mg 24.32	13 Al 26.97	14 Si 28.06	15 P 30.98	16 S 32.066	17 Cl 35.457		18 A 39.944	
4	4		19 K 39.096	20 Ca 40.08	21 Sc 45.10	22 Ti 47.90	23 V 50.95	24 Cr 52.01	25 Mn 54.93	26 Fe 55.85	27 Co 58.94	28 Ni 58.69
5	5		29 Cu 63.54	30 Zn 65.38	31 Ga 69.72	32 Ge 72.60	33 As 74.91	34 Se 78.96	35 Br 79.916		36 Kr 83.7	
6	6		37 Rb 85.48	38 Sr 87.63	39 Y 88.92	40 Zr 91.22	41 Nb 92.91	42 Mo 95.95	43 Tc	44 Ru 101.7	45 Rh 102.91	46 Pd 106.7
7	7		47 Ag 107.880	48 Cd 112.41	49 In 114.76	50 Sn 118.70	51 Sb 121.76	52 Te 127.61	53 I 126.92		54 Xe 131.3	
8	8		55 Cs 132.91	56 Ba 137.36	57-71 Rare earths*	72 Hf 178.6	73 Ta 180.88	74 W 183.92	75 Re 186.31	76 Os 190.2	77 Ir 193.1	78 Pt 195.23
9	9		79 Au 197.2	80 Hg 200.61	81 Tl 204.39	82 Pb 207.21	83 Bi 209.00	84 Po 210	85 At		86 Rn 222	
10	10		87 Fr	88 Ra 226.05	89 Actinide series**							

\* Rare earths: 57 La | 58 Ce | 59 Pr | 60 Nd | 61 Pm | 62 Sm | 63 Eu | 64 Gd | 65 Tb | 66 Dy | 67 Ho | 68 Er | 69 Tm | 70 Yb | 71 Lu | 138.92 | 140.13 | 140.92 | 144.27 | 150.43 | 152.0 | 156.9 | 159.2 | 162.46 | 164.94 | 167.2 | 169.4 | 173.04 | 174.99

\*\* Actinide series: 89 Ac | 90 Th | 91 Pa | 92 U | 93 Np | 94 Pu | 95 Am | 96 Cm | 97 Bk | 98 Cf | 232.12 | 231 | 238.07

(e) In 1894 the startling discovery was made that the atmosphere contained about 1% of a hitherto unnoticed gas, later called *argon* (A), which had escaped previous detection by chemists because of its complete unwillingness to enter into any chemical combination with other elements. Now other such gaseous elements were discovered, forming a family of six inert gases (He, Ne, A, Kr, Xe, Rn). They all could be accommodated together in the periodic scheme by simply adding one vertical column, either interlaced with group VIII or, better, marked as a separate group with group number zero, in accord with the zero valence of each member. This harmonious fit was a considerable triumph for the scheme.

(f) However, Mendeléeff's basic scheme breaks down in a few places. Note that A (argon) and K (potassium) are placed respectively in the 18th and 19th positions as demanded by their chemical properties, whereas on the basis of their respective atomic weights alone ( $A = 39.944$  and  $K = 39.096$ ) one should really transpose the two. Other inversions of this kind can be found in the table, e.g., for elements 52 ( $Te = 127.61$ ) and 53 ( $I = 126.92$ ). Mendeléeff had confidently expected that the atomic weight for Te would on re-determination be found to be lower than that for I. The necessity of an inversion would have been to him not an inconvenience but a catastrophe. In connection with just such a point, he said:

“The laws of nature admit of no exception . . . It was necessary to do one or the other—either to consider the periodic law as completely true, and as forming a new instrument in chemical research, or to refute it. Acknowledging the method of experiment to be the only true one, I myself verified what I could, and gave everyone the possibility of proving or confirming the law, and did not think, like L. Meyer, when writing about the periodic law that ‘it would be rash to change the accepted atomic weights on the basis of so uncertain a starting point.’ In my opinion, the basis offered by the periodic law had to be verified or refuted, and experiment in every case verified it.”

Clearly, he overestimated the *necessity* of the periodic law in every detail, particularly as it had not yet received a physical explanation; and although the anomalous inversions in the sequence of elements are real, their existence did not invalidate the scheme. As so often, the complexity of nature is here greater than any one generalization can accommodate. Yet one must understand that it was just this utter, enthusiastic belief in the existence of a simple

scheme that carried Mendeléeff to his undaunted and far-reaching conclusions.

The usefulness of the periodic law was not confined to the re-determination of some atomic weights and the prediction of properties of new elements. It had provided an inner order in the list of elements that revealed several previously unsuspected analogies among elements. The whole study of inorganic chemistry was thereby revitalized. And perhaps most important of all, just as the Keplerian discoveries of simple harmonies and mathematical order in planetary motions posed the tantalizing problem of how to account for it by a fundamental physical law, so was science now challenged to provide an explanation for the observed regularities of the periodic table in terms of a physical model of the atom. This task, and its solution in our century, so reminiscent of Newton's solution in the analogous case, is to be recounted in Part G. The atom of the last century, however, by and large remained incapable of supplying a deeper insight into the fundamental processes of chemistry. On the other hand, it did achieve startling successes in the hands of the 19th century physicists. And we now turn to pick up that part of the story in the development of the atomic theory.

### Additional Problems

PROBLEM 19-17. In the 1897 English edition of his book *Principles of Chemistry*, we find Mendeléeff's own summary of his original work on the periodic table. Discuss each of the eight points, choosing further illustrations from the material in this chapter:

"The substance of this paper is embraced in the following conclusions: (1) The elements, if arranged according to their atomic weights, exhibit an evident *periodicity* of properties. (2) Elements which are similar as regards their chemical properties have atomic weights which are either of nearly the same value (platinum, iridium, osmium) or which increase regularly (e.g., potassium, rubidium, cesium). (3) The arrangement of the elements or of groups of elements in the order of their atomic weights corresponds with their so-called *valencies*. (4) The elements, which are the most widely distributed in nature, have *small* atomic weights, and all the elements of small atomic weight are characterised by sharply defined properties. They are therefore typical elements. (5) The *magnitude* of the atomic weight determines the character of an element. (6) The discovery of many yet unknown elements may be expected. For instance, elements analogous to aluminum and silicon, whose atomic weights would be between 65 and 75. (7) The atomic weight of an element may sometimes be corrected by aid of a knowledge of those of the adjacent elements.

Thus the combining weight of tellurium must lie between 123 and 126, and cannot be 128. (8) Certain characteristic properties of the elements can be foretold from their atomic weights. The entire periodic law is included in these lines."

PROBLEM 19-18. In the *Handbook of Chemistry and Physics* there is printed, below one of the periodic tables, a plot of valence numbers for the elements *vs.* atomic number. Neglect the *negative* valence numbers and plot (to element 65) a graph of *maximum* valences observed *vs.* atomic weight. What periodicity is found? Is there any physical or chemical significance to this periodicity? Does there have to be any?

### Further References

- T. W. Chalmers, *op. cit.* (Ch. 17). Chapters VI and VII. Very readable.
- J. C. Gregory, *op. cit.* (Ch. 18), to Chapter XII.
- E. J. Holmyard, *Makers of Chemistry*. Oxford: Clarendon Press. Biographies and work of Dalton, Avogadro, Mendeléeff, and many others. See particularly Chapters 50, 52, 53, 55.
- B. Jaffe, *Crucibles*. Cleveland: World Publishing Co. (1930). Sketches the work of Priestley, Lavoisier, Dalton, Avogadro, Mendeléeff, and others.
- W. S. Knickerbocker, *op. cit.* (Ch. 7). Excerpts from Dalton (Ch. XXII) and Avogadro (Ch. XXIV). Also J. W. Knedler, *op. cit.*, pp. 247-272 (Dalton) and pp. 535-567 (Mendeléeff).
- P. Lenard, *op. cit.* (Ch. 9); biographies of Dalton, Gay-Lussac.
- L. K. Nash, *The Atomic-Molecular Theory*. Cambridge, Mass.: Harvard University Press (1950).
- J. R. Partington, *A Short History of Chemistry*. London: Macmillan (1939), particularly Chapters VII, VIII, XV.
- L. Pauling, *General Chemistry*. San Francisco: W. H. Freeman and Co. (1947). Good introductory text.
- D. Q. Posin, *Mendeleyev*. New York: Whittlesey House (1948).
- H. Roscoe, *John Dalton and The Rise of Modern Chemistry*. New York: Macmillan (1895).
- Sir William A. Tilden, *Famous Chemists, The Men and Their Work*. New York: E. P. Dutton (1921). See particularly the chapters on Dalton, Gay-Lussac, Avogadro, Cannizzaro, and Mendeléeff.
- W. P. D. Wightman, *op. cit.* (Ch. 6). Chapters XV and XX.

## CHAPTER 20

### THE KINETIC-MOLECULAR THEORY OF GASES

**20-1 Introduction.** Having watched the establishment of the chemical atom through the labor of Dalton and Avogadro, we finally come, full circle, back to the topic of Democritus and Leucippus, the atomicity of physical matter and the nature of heat. The key figure is an old acquaintance, James Prescott Joule of Manchester, exactly the man best fitted to accomplish that final synthesis by virtue of having Rumford's convictions of a kinetic theory of heat, Bernoulli's gift for a quantitative treatment of the problem, Mayer's certainty in a law of conservation of energy, and, of course, experiments of his own on the equivalence of heat and mechanical work. The final stage in the recognition of the atomic theory of matter came when Joule, in lectures and published papers from 1847 to 1857, resurrected the century-old gas model of Bernoulli, fortifying it carefully with a series of new and convincing arguments and calculations.

To recapitulate, in the intervening century or so, Dalton's atomic theory of chemistry had raised the atomistic idea itself from the qualitative level of speculation; the density and molecular weight of gases were becoming accessible with some accuracy; Rumford and Davy had called the caloric theory in question; Mayer had discussed the idea (in 1842 and 1845) that heat and mechanical energy both obeyed an over-all law of conservation of energy; Joule's own laborious experiments had independently established Mayer's inspired postulate, and had, in fact, yielded a value for the mechanical equivalent of heat; and the old conception that the pressure of gases is to be ascribed to the impact of incessantly moving particles was finding some new champions, as, for example, in the British scientist and physician Herapath. All these were essential stimuli toward a satisfactory extension of Bernoulli's idea.

Joule's paper, *On Matter, Living Force, and Heat* (1847), is largely an argument fashioned after the manner of Rumford: he tries to make plausible the position that a kinetic theory of heat, joined to an atomic theory of matter, can account *qualitatively* quite well for a wide variety of physical phenomena, as shown in the following extract (where we must translate the words *impetus* and *living force* as kinetic energy, *attraction through space* as potential energy).

"A few words may be said . . . with respect to the real nature of heat. The most prevalent opinion, until of late, has been that it is a *substance* possessing, like all other matter, impenetrability and extension. We have, however, shown that heat can be converted into living force and into attraction through space. It is perfectly clear, therefore, that unless matter can be converted into attraction through space, which is too absurd an idea to be entertained for a moment, the hypothesis of heat being a substance must fall to the ground. Heat must therefore consist either of living force or of attraction through space. In the former case we can conceive the constituent particles of heated bodies to be, either in whole or in part, in a state of motion. In the latter we may suppose the particles to be removed by the process of heating, so as to exert attraction through greater space. I am inclined to believe that both of these hypotheses will be found to hold good—that in some instances, particularly in the case of *sensible* heat, or such as is indicated by the thermometer, heat will be found to consist in the living force of the particles of the bodies in which it is induced; whilst in others, particularly in the case of *latent* heat, the phenomena are produced by the separation of particle from particle, so as to cause them to attract one another through a greater space. We may conceive, then, that the communication of heat to a body consists, in fact, in the communication of impetus, or living force, to its particles. It will perhaps appear to some of you somewhat strange that a body apparently quiescent should in reality be the seat of motions of great rapidity; but you will observe that the bodies themselves, considered as wholes, are not supposed to be in motion. The constituent particles, or atoms of the bodies, are supposed to be in motion, without producing a gross motion of the whole mass. These particles, or atoms, being far too small to be seen even by the help of the most powerful microscopes, it is no wonder that we cannot observe their motion. There is therefore reason to suppose that the particles of all bodies, their constituent atoms, are in a state of motion almost too rapid for us to conceive, for the phenomena cannot otherwise be explained. The velocity of the atoms of water, for instance, is at least equal to a mile per second of time [as we shall calculate later]. If, as there is reason to think, some particles are at rest while others are in motion, the velocity of the latter will be proportionally greater. An increase of the velocity . . . of the particles will constitute an increase of temperature, which may be distributed among the neighboring bodies by what is called *conduction*,

that is, on the present hypothesis, by the communication of the increased motion from the particles of one body to those of another. The velocity of the particles being further increased, they will tend to fly from each other . . . overcoming the attraction subsisting between them. This removal of the particles from each other will constitute a new condition of the body—it will enter into the state of fusion or become melted. But, from what we have already stated, you will perceive that, in order to remove the particles violently attracting one another asunder, the expenditure of a certain amount of living force or heat will be required. Hence it is that heat is always absorbed when the state of a body is changed from solid to liquid, or from liquid to gas. . . . When, again, by the application of cold we condense the steam into water, and by a further abstraction of heat we bring the water to the solid condition of ice, we witness the repetition of similar phenomena in the reverse order. . . .”

More quantitative details were given in later papers, where Joule showed how to calculate the speed of gas particles. Taken by itself, this conclusion would hardly seem convincing or of importance, but Joule showed that his model for gases predicted Boyle's law, gave physical meaning to the concept of absolute temperature, and provided a method for interpreting the observed specific heat values for different gases. What is more, other contributors (A. Krönig, 1856; R. Clausius, 1857) soon showed that from Joule's scheme followed a variety of other experimental results, including Gay-Lussac's law and Avogadro's hypothesis.\*

This summary has only hinted at the richness of conclusions and the generality of usefulness of the kinetic-molecular theory which, through the further work of such men as Helmholtz, Maxwell, Boltzmann, and Gibbs, eventually became a major triumph of 19th century physical science. For our purposes it will be well to leave the strictly historical road, and to summarize in modern language some qualitative consequences before we turn to the quantitative side of the concept that heat is a mode of motion of submicroscopic particles.

## 20-2 Some qualitative successes of the kinetic-molecular theory.

(a) Any work done on a substance (e.g., by friction) may be converted into kinetic energy of motion or potential energy between

---

\* A contribution very similar to Joule's was submitted to the Royal Society by J. J. Waterston in 1845, but had been filed away with the comment that it was without sufficient merit for consideration, and lay forgotten for nearly 50 years.



the atoms or molecules. In gases, the increase in kinetic energy of the particles means greater random thermal motion; in solids, the energy may be stored in increased vibration of the crowded atoms; and in a liquid, both thermal motion and thermal vibration play their parts. Heat conduction from a hot to a cold body is thus to be thought of as the transfer of energy of motion from one set of molecules to the other at the interface, the surface of contact. Expansion of a solid or liquid on heating is then the consequence of increased agitation of individual molecules, and their resulting greater separation through more powerful mutual collisions. Therefore, if a glass-in-mercury thermometer is inserted in a vessel containing a gas, and if now work is done on the gas (by compression or friction), the thermometer indicates a rise to a higher temperature reading on the scale because (i) the work done on the gas serves to increase the kinetic energy of the gas molecules, which in turn means (ii) stronger, more frequent collisions of these molecules with the mercury-containing glass bulb, (iii) a transfer of some of this energy to the mercury, and (iv) ultimately an expansion of the mercury column as the increased agitation of mercury molecules results in their greater mutual separation.

(b) Our own sensations of heat or cold by touching are not so simply explained by this or any other physical theory, but evidently we can neither feel nor see directly the individual agitation of the molecules, for their dimensions are a thousand times smaller than the wavelength of light or the sensitive receptors of our skin. Yet, if we consider a small dust particle or bacterium suspended in still air or on the surface of a liquid and bombarded from all sides by swarms of invisible but highly agitated molecules of the surrounding gas or liquid, we understand why under high-power magnification the particle is seen to dance in perpetual, irregular motion, as though alive. In fact, when first observed on tiny grains of plant pollen in 1827 by the English botanist Robert Brown, this phenomenon (now called *Brownian movement*) was suspected of being evidence of "vitality" (see Fig. 20-1).

(c) The extract from Joule's paper has shown two other qualitative successes of the kinetic-molecular theory: an explanation of the difference between the three states of matter—solid, liquid, and gaseous—in terms of the relative agitation of the molecules, and a plausible reason for the existence of latent heats of fusion and evaporation. The same fundamental ideas also help to make clear several other simple phenomena, for example, (i) the diffusion of one gas through another (one type of thermally agitated molecules working

their way through a space filled with another type of molecules), or (ii) the immediate expansion of a gas to fill any vessel no matter how large (since the mutual attractive forces between gas molecules are supposedly negligible and the velocities of random motion very high), and (iii) the very existence of pressure exerted by a gas against its container (understandable as the cooperative effect of billions of collisions per second by all the agitated gas molecules with the walls of the container).

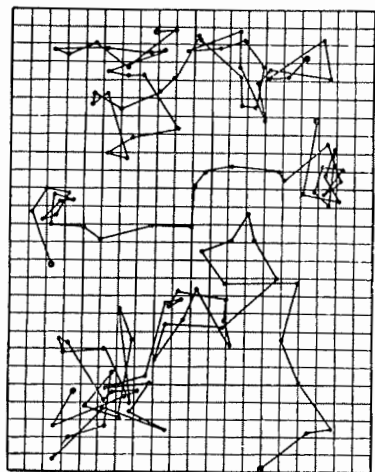


FIG. 20-1. Brownian movement. The positions of three grains on water, as seen through the microscope at 30-second intervals. (After J. Perrin.)

### 20-3 Model of a gas, and assumptions in the kinetic theory.

Turning now to quantitative achievements of the kinetic-molecular theory (and dropping "molecular" from the phrase, in conformity with custom), we shall concentrate mainly on its applications to *gases*, this being most illustrative, mathematically easiest, and, in fact, physically most developed at the present time. We shall follow a simplified derivation, in which our problem reduces to this: Let us construct a plausible *model of a gas* which will yield as a logical consequence (thus "explain") the known laws of the behavior of gases, and which if possible will lead us to look for new verifiable effects. The model and the underlying hypotheses leading to the model together form the *theory* which is useful to the extent that it explains previously known laws and generates new results.

In constructing the model of a gas, we now carefully examine all clues. Some may be well-known experimental facts about matter, others may be plausible but hitherto unchecked assumptions, and still others may be bold new hypotheses which are justifiable *a posteriori* if they lead to verifiable results. Let us summarize the ten major facts or assumptions from which we must build up our initial model, with the experimental evidence for each, where meaningful:

(a) *The general gas law.* The experimental facts about a given

sample of a relatively rarefied and stable gas are, as we saw, well summarized in the equation  $PV/T = r$ .\*

(b) *Molecular hypothesis.* One of the fundamental assumptions of our gas picture is, of course, the idea that gases are made of tiny submicroscopic particles—atoms or groups of atoms called molecules. The best experimental evidence originally was chemical—the success of Dalton's atomic theory, particularly in dealing with the laws of combining proportions. For our purposes we shall at first discuss only pure single gases made of identical building blocks (one type of molecule or one type of atom).

(c) *The equation of state.* Another experimentally supported concept, which we may either build into our model at the start or else must demand as one of its derivable consequences, is Avogadro's hypothesis, which, simply stated, requires that all true gases at a given pressure, volume, and temperature have the same number of molecules. Having defined Avogadro's number  $N$  as the number of molecules per mole of substance ( $6.02 \times 10^{23}$  molecules per mole of any substance), we can reformulate the general gas law, which reads

$$\frac{PV}{T} = r \quad (20-1)$$

and which has the inconvenient character that the value of  $r$  has to be determined experimentally for every particular sample of every gas under investigation ( $r$  changes both with the mass and with the chemical nature of the gas). Specifically, experiments with samples of the same gas but with different total masses  $m$  would reveal that

$$\frac{PV}{T} \propto m.$$

Furthermore, choosing gas samples of equal mass  $m$  but with different chemical constitution (i.e., different molecular weight  $M$ ) would yield the relation

$$\frac{PV}{T} \propto \frac{1}{M}.$$

There being no other factors that need to be taken into account, we can write

\* At extreme conditions of pressure and temperature this simple equation is no longer sufficiently accurate but can be replaced by a better approximation which is theoretically nicely defensible. Ideal, perfect, or true gases, by definition, are gases that behave accurately according to  $PV/T = \text{constant}$ —which fortunately applies to most actual gases far from such extreme conditions.

$$\frac{PV}{T} \propto \frac{m}{M} \quad \text{or} \quad \frac{PV}{T} = \frac{m}{M} \times (\text{a constant } R), \quad (20-2)$$

where  $R$  is indeed a factor equal for every sample of every gas, and therefore may be called the *universal gas constant*. You will also note that  $(m/M)$ , being the ratio of the mass of a sample and its molecular weight, is nothing else than  $n$ , the number of moles of the gas sample at hand. Therefore we now write

$$PV = nRT. \quad (20-3)$$

This most important equation, which is essentially a summary of experimentally observed facts, is the so-called *equation of state* for (perfect, ideal) gases. Since  $R$  is a *universal constant*, the equation  $R = PV/nT$  should give us a numerical value for  $R$  if we know any one set of values of  $P$ ,  $V$ ,  $n$ , and  $T$  which all gases take on simultaneously. And indeed we do: at  $T = 273^\circ\text{K}$  and  $P = 1.013 \times 10^6$  dynes/cm<sup>2</sup>, all gases have, according to Avogadro's hypothesis, the same volume per mole ( $V/n = 2.24 \times 10^4$  cm<sup>3</sup>/mole). Thus  $R = 8.31 \times 10^7$  (ergs/mole)/degree (check the value and the units).

The new form of the gas law has enormous practical advantages over Eq. (20-1). Now we can predict the physical behavior of a quantity of gas if only we know its chemical consistency and its mass! This is evident in the following example, which could not have been solved directly before Eq. (20-3) was formulated.

*Example 1.* What pressure will 10 gm of helium exert if contained in a 500 cm<sup>3</sup> cylinder at room temperature (21°C)?

*Solution:* The relative molecular weight of He is approximately 4 (see Table 19-1). One mole would be 4 gm,  $x$  gm corresponds to  $x/4$  moles, therefore  $n = 10/4 = 2.5$  moles.

$$P = \frac{nRT}{V} = \frac{[2.5 \text{ moles}][8.31 \times 10^7 \text{ (ergs/mole)/degree}] \times [294^\circ\text{K}]}{500 \text{ cm}^3}$$

$$= 1.22 \times 10^9 \text{ dynes/cm}^2,$$

which is over 1000 atmospheres (and therefore probably too much pressure for this container).

PROBLEM 20-1. (a) A 50 cm<sup>3</sup> cylinder is safe only up to 100 atmospheres. Find at what temperature 3 gm of oxygen gas will burst the cylinder. (b) Find the number of molecules of oxygen that must be present in a vessel of 10 cm<sup>3</sup> capacity at 20°C to produce a pressure of 50 atm.

To repeat:  $PV = nRT$  is an experimentally verified equation for ideal gases, and therefore it must either appear axiomatically in our gas model, or else must follow as a consequence of the model.

(d) *Size of molecules.* The model of the gas, which we are piecing together, must include the provision that the individual molecules are extremely small, perhaps of the order of  $10^{-8}$  to  $10^{-7}$  cm. The experimental evidences for this point are numerous, varied, and conclusive. For example, vapors and gases occupy volumes that may be thousands of times larger than the liquids from which they can be obtained and, conversely, gases may be compressed to minute fractions of their original volume. They diffuse through invisibly small pores in unglazed porcelain. They combine in exact accord with the laws of chemical proportions even in extremely minute quantities. Even as clumsy and large molecules as those of oil (glycerin trioleate,  $C_{57}H_{104}O_6$ ) cannot exceed  $10^{-7}$  cm in diameter because, by measurement, that is the approximate thickness of oil films spread very thinly on water.

We conclude therefore that ordinarily we may neglect the actual volume of all molecules in comparison with the total volume which the gas occupies by virtue of its unceasing motion, and indeed we might at the start consider them as point masses, or at any rate as very tiny, smooth spheres.

(e) *Number of molecules.* By the same evidence as in (d) we must suppose that there are enormous swarms of molecules in every cubic centimeter of ordinary gases; in fact, Avogadro's number expresses this point quantitatively. Even in the best vacuum ordinarily obtained in the laboratory (about  $10^{-10}$  atm) there are at ordinary temperatures still about  $10^9$  molecules/cm<sup>3</sup>.

(f) *Mean free path.* As the enormous expansion of a gas obtained from a liquid indicated, the average distances between molecules are correspondingly huge compared with their own size. On the average, a molecule may travel for distances hundreds or thousands of times its own dimensions before encountering another molecule; these distances, measured indirectly, are of the order of  $10^{-5}$  cm for ordinary densities, and are referred to as *mean free path*.

(g) *Forces between molecules.* Although the mean separation of molecules may be large by atomic dimensions, it is so small an absolute quantity that we may wonder about the mutual forces between gas molecules (which we guess to be strong in the liquid and solid states because of the large latent heats).

First we may suspect conventional gravitational forces, but they

are trivial, as a quick calculation will show. The masses of, say, oxygen molecules, are calculated from Avogadro's postulate. Since 32 gm of  $O_2$  contain over  $6 \times 10^{23}$  molecules, each molecule weighs about  $32/(6 \times 10^{23})$  or about  $5 \times 10^{-23}$  gm. Two such molecules at about  $10^{-5}$  cm apart exert a mutual gravitational force of

$$F = G \frac{m_1 m_2}{d^2}, \quad \text{or} \quad F \doteq 6.7 \times 10^{-8} \times \frac{(5 \times 10^{-23})^2}{(10^{-5})^2} \text{ dyne,}$$

about  $3 \times 10^{-27}$  dyne, evidently a completely negligible quantity. (So also is the weight of each molecule by itself for all our purposes.)

But there may be among these particles forces other than gravitational—perhaps electrical ones. Although we may not readily imagine exactly what or how strong they are, we can make an experiment to see whether or not they exist. If they do, then a gas under pressure which is allowed to expand into a big *evacuated* vessel, on separating from its fellow molecules, will invoke such intermolecular forces to do work. If the forces are attractions, then work is done by the molecules on separating and since, according to the fundamental assumptions, the required energy for this process can come only from heat energy, the gas ought to cool itself on expanding into a vacuum. Conversely, if the intermolecular forces are repulsions, the total mass of the gas should heat up on expansion. The experimental fact observed previously by Gay-Lussac (see Section 17-8), is, however, that on expansion into a vacuum only a very small amount of net cooling or heating occurs in the whole mass of the actual gases,\* hence no significant intermolecular forces exist except during actual collisions, which presumably take very little time and on the average occupy a negligible span for each molecule compared with its whole history. Incidentally, you will recall that this is exactly the experiment which most directly refutes the old *static* model of mutually repelling gas molecules.

Here we may digress for a moment: even if it be true that the molecules do not exert significant forces on one another while in gaseous form and widely separated, we must still expect that when they come closer together, as in liquids and solids, such intermolecular forces become important. We must postulate that slight increases in the intermolecular separation distances call into action a certain amount of mutual *attraction* in the case of liquids and solids, for otherwise we cannot explain *cohesion*, the resistance offered to efforts

\* For an excerpt of Joule's own original experiments, see Magie, *op. cit.*, pages 172-174.

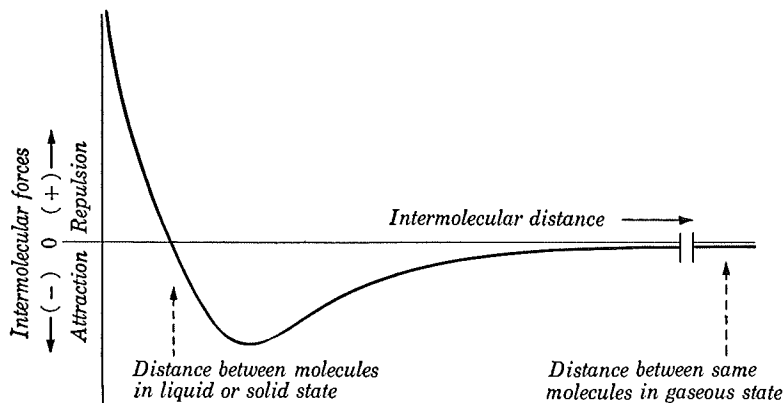


FIG. 20-2. Qualitative graph of the force between two molecules *versus* the intermolecular distance.

tending to separate the material into parts. There also must be the converse effect, mutual *repulsion* on decreasing the intermolecular distance, otherwise we cannot account for the large resistance of liquids and solids to the slightest compressions. One may assemble all this information about intermolecular forces in one qualitative graph of distance *vs.* force (where *positive* forces represent repulsion among molecules, *negative* ones attraction; see Fig. 20-2).

PROBLEM 20-2. Justify the indicated position in Fig. 20-2 for the intermolecular distances in solids and gases.

(h) *Thermal agitation.* Strongly implied in the developing model, in the fundamental assumptions of the kinetic theory, and in the observations on Brownian movement and diffusion, is the requirement of *rapid* motion of the gas molecules at usual and high temperature. The picture of the gas then expands to the view of moving particles, passing through relatively "long" stretches of straight motion, and colliding briefly but many million times per second with other molecules or with the walls of the container (where each collision contributes to the maintenance of pressure). So brief are these mutual collisions, so random the motions, that we shall in our analysis neglect the effect of such encounters entirely. We shall also see that thermal agitation, as implied by the observation of Brownian movement, must be omnipresent. All assemblies of atoms (if above  $0^{\circ}\text{K}$ ) exhibit such agitation, not only in gases and liquids but even in solids, where it is expressed by a small-scale vibration among neighboring atoms.

(i) *Nature of collisions; Newton's laws apply.* Now that we have allowed the molecules to undergo collisions, we must consider the mechanical laws governing such events. Naturally, we turn to Newton's mechanics, although we have no direct evidence that microscopic worlds obey these laws, which had developed from experiments with large-scale objects. But although we may have to keep our eyes open to retract this dangerous extrapolation if facts demand it, we are more or less forced to start with the assumption that Newtonian laws hold among molecules, for we have no *other* mechanics available to us at this point. In his pioneering work on this subject, James Clerk Maxwell discusses the relation between the atomistic model and the forces of interaction as follows:

“Instead of saying that the particles are hard, spherical, and elastic, we may if we please say that the particles are centers of force, of which the action is insensible except at a certain small distance, when it suddenly appears as a repulsive force of very great intensity. It is evident that either assumption will lead to the same results.”

(j) *Conservation of kinetic energy.* We must also assume that the molecular collisions with one another and with the container are perfectly elastic, i.e., that on the average the kinetic energy (KE) of a molecule before the collision is the same as immediately afterward; only *during* the brief collision will the KE pass into potential energy of elastic distortion, presumably just as in the case of an elastic ball bouncing off a wall.

The reasons for this simplifying assumption are evident: If in the long run the collision were not perfectly elastic, the KE of each molecule would gradually disappear even in a gas filling a heat-insulated container completely left to itself; with a diminishing total KE goes a reduced pressure, until eventually the speeds are so low that the feeble gravitational and other attractive forces induce liquefaction or solidification of the gas.

Nothing like that is observed at all. Nor can we at this point readily imagine what could happen to KE lost in a hypothetical inelastic collision of a molecule. Of course some energy could be—and sometimes is—somehow stored as potential energy, as in a change of position or increased vibration of the parts of a molecule with respect to one another or even in rotation of the molecule. But this is not an unlimited reservoir for energy, and in any case we should



try out the simpler hypothesis of perfectly elastic behavior.\* The previously mentioned suggestion that the molecules be thought of as tiny, round, smooth spheres is simply an expression of this requirement of conservation of kinetic energy.

**20-4 The derivation of the pressure formula.** We now have a fine mental picture or *model* of a typical ideal gas. It is overpoweringly tempting to visualize it in terms of a three-dimensional assembly of agitated microscopic billiard balls; therefore a warning is in order to be prepared for a breakdown of the analogy where it is carried beyond the evidence and for new assumptions to be introduced where necessary, and to be ready for surprising consequences of our model and for behaviors quite beyond anything an assembly of billiard balls would ordinarily dream of doing.

As an initial test of the validity of our model we now choose any well-established *quantitative* law of gases and try to deduce it from the model. Following Joule's choice, we decide on the general gas law [in its modern form, Eq. (20-3)],  $PV = nRT$ . If we can derive this equation, we should be completely satisfied. But actually we cannot do this directly, and instead we shall obtain from our analysis another equation, of the form  $PV = XYZ$  (where  $XYZ$  stands for some factors such as the mass and velocity of the molecules). The proof that our model is valid then depends on showing that  $XYZ = nRT$ , or that the equation  $PV = XYZ$  actually yields experimentally verifiable consequences. This we shall do, with exciting secondary results.

To derive a relationship between the pressure, volume, and other variables of a gas, consider a strong and rigid cubical container with sides of length  $l$ ,\*\* enclosing the molecules of a pure ideal gas (Fig. 20-3). Each molecule has the same mass,  $m_0$ , and moves rapidly back and forth in the box, frequently colliding with the six walls, in accordance with the gas model which we have agreed on.

---

\* On the other hand, no large-scale body, not even a sphere of the best steel, shows perfectly elastic collisions; there are always some losses of mechanical energy through internal friction and a consequent rise in temperature through increased internal molecular agitation. But what could be the meaning of frictional heat losses in the collisions of the molecules themselves? Note that we have here a good example of the breakdown of a concept gained from large-scale experience when applied to the realm of the submicroscopic.

\*\* If this first simplifying assumption seems too restrictive, consider that any real noncubical container may be thought of as a set of many such very tiny cubical subspaces.

(a) First imagine that there is only a single molecule rattling around in the box. For simplicity, we also assume that its speed  $v$  (and therefore its kinetic energy) does not change numerically except during actual collision, which, however, takes only a negligibly short time. Every time the molecule collides with one flat and rigid side of the box, say  $A$  (shaded in Fig. 20-3), it is reflected again with the same speed; evidently the component of the velocity perpendicular to side  $A$  is simply reversed, and the other two components of the velocity are not affected at all by that collision.

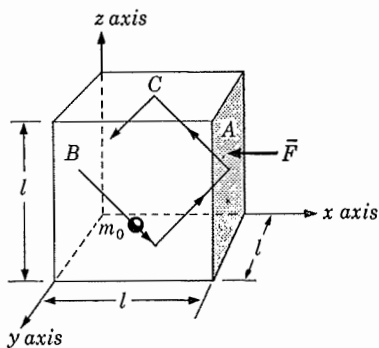


FIG. 20-3.

(b) If  $v_x$  is the velocity component of the molecule in the  $x$  direction, and  $t$  is the time which the molecule needs to travel the  $l$  cm from side  $B$  (opposite  $A$ ) to side  $A$ , then

$$v_x = \frac{l}{t}, \quad \text{and} \quad t = \frac{l}{v_x}.$$

Then  $\tau$ , the time needed for a complete round trip from one collision with side  $A$  to the next collision with that side, is  $2t$ . Thus we write:

Time elapsed per collision (with side  $A$ , for one molecule)  $\equiv \tau = 2l/v_x$ .

If  $\tau$  is the number of seconds per one collision, the reciprocal quantity  $1/\tau$  is the number of collisions per second, given by  $1/\tau = v_x/2l$ .

(c) During each of the  $v_x/2l$  collisions occurring every second between the molecule of mass  $m_0$  and the rigid side  $A$ , the velocity component  $v_x$  is changed from positive (to the right in Fig. 20-3) to negative (to the left), a total change of  $2v_x$ . The change of momentum of the molecule (being the mass times the change of velocity)  $= 2m_0v_x$  for each collision, but during one second, the sum of all such momentum changes is

$$(2m_0v_x) \times \left(\frac{v_x}{2l}\right),$$

i.e., (momentum change/collision)  $\times$  (collisions/sec). Of course,

$$(2m_0v_x) \times \left(\frac{v_x}{2l}\right) = \left(\frac{m_0v_x^2}{l}\right),$$

and this still applies for only a single molecule. But we now consider not one but the actual number,  $N_m$ , of molecules in this box. Their total change of momentum per second at side  $A$  is

$$N_m \left( \frac{m_0 v_x^2}{l} \right),$$

if they all ordinarily have the same speed. If they do not (and it would be too much of a limitation to put in such an improbable restriction) we need to do no more at this time than to write  $\overline{v_x^2}$  instead of  $v_x^2$  in the last expression, where  $\overline{v_x^2}$  means the *mean* (straight average) value of the sum of the square of the  $x$  components of velocity for all  $N_m$  molecules present; i.e.,

$$\overline{v_x^2} \equiv \frac{[(v_x)_1^2 + (v_x)_2^2 + (v_x)_3^2 + \cdots + (v_x)_{N_m}^2]}{N_m}. \quad (20-4)$$

This presupposes nothing about the way actual values of  $v_x$  are distributed among the  $N_m$  molecules, a question we may investigate later.

Thus we now write: the total change of momentum per second of all  $N_m$  molecules at side  $A$  is equal to

$$\frac{N_m m_0 \overline{v_x^2}}{l}.$$

(Watch how this expression is growing and changing meaning in the development.)

(d) By Newton's laws, the change of momentum = average net force  $\overline{F}$  exerted, multiplied by the *time* during which the change of momentum occurs. Consequently, change of momentum *per second* for the molecules =  $\overline{F}$ , numerically; or, the average force exerted on the molecules by the side  $A$  as they rebound is numerically given by

$$\overline{F} = \frac{N_m m_0 \overline{v_x^2}}{l}. \quad (20-5)$$

Since the hail of molecules against the wall is pretty thick, this force is what might be observed steadily if the wall or side  $A$  were a piston supported only by means of a compression-type spring balance against the blows of the molecules. By the law of action and reaction it is, of course, also numerically equal to the average (perpendicular) force with which the molecules push on side  $A$ .

(e) Since we are not after the force exerted on the wall but after the *pressure*, we recall that pressure  $P$  = perpendicular force/area.

If we divide both sides of Eq. (20-5) by  $l^2$ , which is the area of side  $A$  (and also the area of every one of the other five sides), we now have

$$\frac{\bar{F}}{l^2} (\equiv P \text{ on side } A) = \frac{N_m m_0 \bar{v}_x^2}{l^3}.$$

But  $l^3$  is the volume  $V$  of the cubical container, hence

$$P = \frac{N_m m_0 \bar{v}_x^2}{V} \quad \text{or} \quad PV = N_m m_0 \bar{v}_x^2. \quad (20-6)$$

This is a very promising development in our effort to derive the relationship between  $P$ ,  $V$ , and other measurable quantities; all that remains to be done is to prove that the right side of Eq. (20-6) is equivalent to  $nRT$ , but that will not be the easiest part of our work!

(f) First of all, the expression  $\bar{v}_x^2$  in Eq. (20-6) can be replaced by  $\bar{v}^2/3$ ; the argument is rather simple. From Eq. (20-6),

$$\bar{v}_x^2 = \frac{PV}{N_m m_0},$$

where  $P$  is the pressure on face  $A$ , and  $V$ ,  $N_m$ , and  $m_0$  are constants for this box and this gas. If we had chosen to pay attention to one of the other faces, say the top face  $C$  which stands perpendicular to the  $z$  axis, we should have obtained  $\bar{v}_z^2 = \frac{PV}{N_m m_0}$ . The pressure  $P$  at  $C$  is no different than on  $A$  or on any other face, and  $V$ ,  $N_m$ , and  $m_0$  are still the same for each case, therefore

$$\bar{v}_x^2 = \bar{v}_y^2 = \bar{v}_z^2,$$

i.e., we assume that there is *no preference* among the molecules for motion along any one of the three axes. By Pythagoras' theorem, extended to a vector  $v$  in three dimensions,  $v^2 = v_x^2 + v_y^2 + v_z^2$ , therefore we may write  $\bar{v}^2 = \bar{v}_x^2 + \bar{v}_y^2 + \bar{v}_z^2$ , or in our case,  $\bar{v}^2 = 3\bar{v}_x^2 = 3\bar{v}_y^2 = 3\bar{v}_z^2$ . Therefore  $\bar{v}_x^2 = \frac{1}{3}\bar{v}^2$ , and Eq. (20-6) can be rewritten:

$$PV = \frac{1}{3} N_m m_0 \bar{v}^2. \quad (20-7)$$

Note that  $\frac{1}{2} m_0 \bar{v}^2$  would be the average kinetic energy of translation ( $\overline{\text{KE}}_{\text{trans}}$ ) *per molecule*, and  $N_m (\frac{1}{2} m_0 \bar{v}^2)$  is the *total*  $\text{KE}_{\text{trans}}$  for all  $N_m$  molecules in the sample of gas in our box. Therefore it becomes significant to recast Eq. (20-7):

$$PV = \frac{2}{3} N_m \frac{m_0 \bar{v}^2}{2} = \frac{2}{3} (\text{total } \text{KE}_{\text{trans}} \text{ of gas}). \quad (20-8)$$

In words, the product of pressure and volume of a gas is  $\frac{2}{3}$  the total kinetic energy of translation of the molecules.

This is an astonishing consequence of our model. In a qualitative way we supposed all the time that the pressure of the gas against the container was due to the incessant blows of highly agitated gas molecules, but we had no idea that so simple and beautiful a relation between the pressure and the energy of the gas molecules could exist. Equation (20-8) is the so-called pressure equation for an ideal gas. We may be chagrined that we did not obtain  $PV = nRT$  by our calculations; but we have the promised relation  $PV = XYZ$ , and so the validity of our model can still be tested by examining whether the proposition  $XYZ = nRT$  yields verifiable consequences. Moreover, if the factor  $\bar{v}^2$  depends only on the temperature of the gas (to be proved), then Eq. (20-7) represents Boyle's law:  $PV = \text{constant}$  at given  $T$ . This is, of course, most gratifying.

**20-5 Consequences and verification of the kinetic theory.** (a) First let us see whether Eq. (20-8) by itself is reasonable from the point of view of *direct experimental verification*.

It follows from Eq. (20-7) or (20-8) that  $\bar{v}^2 = (3PV/N_m m_0)$ . But  $N_m m_0 = \text{total mass of gas actually present}$ , and  $(N_m m_0/V) = \text{mass density } \rho \text{ of the gas (gm/cm}^3\text{)}$ .

Therefore,  $\bar{v}^2 = 3P/\rho$  and  $\sqrt{\bar{v}^2} = \sqrt{3P/\rho}$ . The expression  $\sqrt{\bar{v}^2}$ , for which we shall use the symbol  $c$ , is the square root of the mean (straight average) of the squares of the various velocities—in brief, the “root-mean-square” value of the molecular velocities (or rms velocity). It is mathematically not the same thing as a straight mean value  $\bar{v}$ , but, in fact, we find that the value of  $\sqrt{\bar{v}^2}$  differs only a few per cent from the straight average\* in actual gases, or  $c \doteq \bar{v}$ .

Now the prediction following from our model is that the rms velocity  $c$ , or nearly enough the average velocity  $\bar{v}$  of the molecules, should be  $= \sqrt{3P/\rho}$ . Substituting representative values for air

\* To appreciate the difference between  $\bar{v}$  and  $c$ , note that if there were only three molecules with speeds  $v_1 = 5$  units,  $v_2 = 10$ , and  $v_3 = 15$ ,

$$\bar{v} = \frac{5 + 10 + 15}{3} = 10 \text{ units,}$$

$$c \equiv \sqrt{\bar{v}^2} = \sqrt{\frac{5^2 + 10^2 + 15^2}{3}} = 10.8 \text{ units.}$$

This is a difference of 8%.

at one atmosphere pressure ( $1.01 \times 10^6$  dynes/cm<sup>2</sup>) and 0°C (where  $\rho = 0.00129$  gm/cm<sup>3</sup>) we find that the rms velocity should be

$$\sqrt{\frac{3 \times 1.01 \times 10^6}{0.00129}} = 4.85 \times 10^4 \text{ cm/sec,}$$

nearly 500 meters/sec, or about  $\frac{1}{2}$  mile/sec—comparable to the speed of a slow bullet!\* (We may now confirm Joule's conclusions that the average velocity of the molecules of water vapor, say at 100°C, is in excess of 1 mile/sec.) This is an enormous speed compared with the motion of everyday objects, but it is of the same order as the speed of propagation of *sound* at that pressure and temperature ( $3.31 \times 10^4$  cm/sec). The rms velocities of molecules in other gaseous materials at that pressure and temperature have, by calculation, similarly high values, ranging from  $18 \times 10^4$  cm/sec for hydrogen gas to  $1.7 \times 10^4$  cm/sec for mercury vapor, and in each case the velocity of sound is not much less. But this is just what we should expect in terms of our model of a gas, for the obvious way to think of the propagation of sound waves is to visualize it as a directional motion of the molecules superposed on their random chaotic motion, so that the energy of the sound wave is carried as kinetic energy from one gas molecule to its next near neighbors with which it collides. The molecules themselves, despite their high speeds, do not move very far during any appreciable interval, being contained within a relatively small volume by multiple collisions, around a billion per second, with their neighbors—but the energy of the sound wave is communicated from one to the next molecule with that high speed.\*\* As to the numerical values involved here, we do not really expect that the speed of sound should be exactly as great as  $c$  [why not?]; in fact the numerical discrepancy is exactly predictable from the complete theory of sound propagation.

**PROBLEM 20-3.** Dalton, who first of all was a meteorologist, discovered experimentally a law concerning gas pressures, which we call Dalton's Law of Partial Pressures, and which may be stated thus: if several (e.g., three)

\* Bernoulli might have carried his own work to this point if he had had good measurements of gas densities.

\*\* Incidentally, this picture of molecules being restrained by mutual collisions despite their individual large speeds explains why the scent of a bottle of perfume (being nothing but a swarm of vapor molecules from that liquid) will take a relatively long time to travel from one corner of a room to the other; this realization disposed of one of the initial strong objections to the kinetic theory.

gases are mixed in one container at a given temperature, then the total pressure  $P$  exerted by all of them together is the sum of the individual pressure (e.g.,  $P = P_1 + P_2 + P_3$ ) which each would exert if it were alone in that container at that temperature. Prove that our kinetic theory model is in harmony with this law. (*Hint*: Accept the general gas law as applicable even to gas mixtures. From Eq. (20-8),  $P = (\frac{2}{3})$  total  $\text{KE}_{\text{trans}}/V$ , and note that kinetic energies add by simple scalar addition.)

(b) Returning to the main line of the argument, the fact that the theoretical result of our model ( $PV = \frac{1}{3} N_m m_0 \bar{v}^2$ ) must coincide with the experimental facts about gases ( $PV = nRT$ ), we proceed to look for derivable and *verifiable* consequences of the implication that the right sides of both equations are equal, that

$$nRT = \frac{1}{3} N_m m_0 \bar{v}^2. \quad (20-9)$$

This can be rewritten:

$$T = \frac{2}{3} \left( \frac{N_m}{nR} \right) \left( \frac{m_0 \bar{v}^2}{2} \right). \quad (20-10)$$

That is, *the temperature of a gas (in °K) is proportional to the mean kinetic energy of translation per molecule for the gas*—a most important result, and a very fruitful interpretation of the concept of temperature.

Furthermore, note that the factor  $\frac{2}{3} (N_m/nR)$  is a universal constant for *all* gases. The proof is as follows:

$$\frac{N_m}{n} = \frac{\text{number of molecules in given sample}}{\text{moles of gas in sample}}.$$

But the number of molecules per mole is, by definition, Avogadro's number,  $N$ , a universal constant. Therefore

$$\frac{2}{3} \left( \frac{N_m}{nR} \right) = \frac{2}{3} \frac{N}{R}.$$

But  $R$  is also a universal constant, and so it is convenient to write the last expression as

$$\frac{2}{3} \frac{1}{k}, \text{ where } k = \frac{R}{N} = \frac{8.31 \times 10^7 \text{ (ergs/mole)/degree}}{6.02 \times 10^{23} \text{ molecules/mole}} \\ = 1.38 \times 10^{-16} \text{ (erg/degree)/molecule.}$$

(The universal constant  $k$  is of greatest significance in many branches of physics, and is called *Boltzmann's constant*, named in honor of the

great 19th century Austrian physicist whose work paved the way to many advances in this field.)

We now rewrite Eq. (20-10):

$$T = \left( \frac{2}{3k} \right) \frac{m_0 \bar{v}^2}{2} = (\text{constant}) \times (\overline{\text{KE}}_{\text{trans}} \text{ per molecule}). \quad (20-11)$$

In words (since  $k$  is truly a universal constant), *the temperature of a gas is directly proportional to the mean kinetic energy of translation per molecule for any gas.*

(c) Consider one of the many consequences of this last statement. At the same temperature  $T$ , for two different types of gases (with values of  $m_1$  and  $v_1$  for the one,  $m_2$  and  $v_2$  for the other) our theory requires that

$$T = \frac{2}{3k} \frac{m_1 \bar{v}_1^2}{2} \quad \text{and} \quad T = \frac{2}{3k} \frac{m_2 \bar{v}_2^2}{2}, \quad \text{or} \quad \frac{m_1 \bar{v}_1^2}{2} = \frac{m_2 \bar{v}_2^2}{2}. \quad (20-12)$$

Hence, at a given temperature, the molecules of any gas whatever have the same  $\overline{\text{KE}}_{\text{trans}}$  as any other gas, or

$$\frac{\bar{v}_1^2}{v_2^2} = \frac{m_2}{m_1}, \quad \text{and} \quad \frac{\sqrt{\bar{v}_1^2}}{\sqrt{\bar{v}_2^2}} \equiv \frac{c_1}{c_2} = \sqrt{\frac{m_2}{m_1}}. \quad (20-13)$$

In words, *the ratio of rms speeds of molecules of two gases is equal to the square root of the inverse ratio of their masses.*

This conclusion is susceptible to immediate experimental test. If we take two different gases at equal temperatures, each in its own tight rigid container at equal pressures, and then open in each container an about equally tiny hole toward a vacuum for the molecules to escape, the gas with the lighter molecular mass, having a relatively higher rms speed, should escape faster than the heavier gas with lower rms speed. This experiment in *effusion* is in principle quite possible, and although the experiment itself obviously would be difficult to perform, the result does, in fact, prove Eq. (20-13) to be correct.

**PROBLEM 20-4.** The mass of each molecule of hydrogen gas is to that of a nitrogen gas molecule as 2:28 (approximately). What are the relative speeds of effusion for two such gases under equal conditions of temperature, etc.?

Another proof of Eq. (20-13) is to be found in the *diffusion* of two different gases trapped in one container with porous walls toward a vacuum; the lighter gas will again escape faster than the heavier one.



This had been established empirically by Thomas Graham in 1830, and now was explainable in terms of a theory.

This process of diffusion is the basis of one type of large-scale separation of the light and heavy isotopes of uranium, the fateful metal. Specifically, uranium (chemical symbol U), as found in nature, consists primarily of an intimate mixture of two chemically equivalent but not equally heavy types of atoms, the isotope U-238 (i.e., relative molecular weight approximately 238, making up 99.3% of natural pure uranium) and the more potent but rarer isotope U-235 (molecular weight approximately 235, abundance about 0.7%). Small traces of a third type of uranium atom, U-234, are inevitably also present, but are negligible now. The important problem of "enriching" the supply of natural uranium, of increasing the relative abundance of U-235 with the help of our kinetic theory, is discussed in the famous *Smyth Report* of 1945.\* For the complete understanding of the following extract, note that the ratio  $c_1/c_2$  in Eq. (20-13) corresponds also to the ratio of the *number* of molecules with these respective speeds which will have found their way through a porous barrier of a vessel after a short time interval. This ratio  $r_1/c_2$  is therefore also called "the ideal separation factor," symbolized by  $\alpha$ . The *Smyth Report* gives these details (pages 115-116):

"As long ago as 1896 Lord Rayleigh showed that a mixture of two gases of different atomic weight could be partly separated by allowing some of it to diffuse through a porous barrier into an evacuated space. Because of their higher average speed the molecules of the light gas diffuse through the barrier faster, so that the gas which has passed through the barrier (i.e., the 'diffusate') is enriched in the lighter constituent, and the residual gas (which has not passed through the barrier) is impoverished in the lighter constituent. The gas most highly enriched in the lighter constituent is the so-called 'instantaneous diffusate'; it is the part that diffuses before the impoverishment of the residue has become appreciable. If the diffusion process is continued until nearly all the gas has passed through the barrier, the average enrichment of the diffusate naturally diminishes. . . . On the assumption that the diffusion rates are inversely proportional to the square roots of the molecular weights, the separation factor

---

\* H. D. Smyth, "A General Account of the Development of Methods of Using Atomic Energy for Military Purposes Under the Auspices of the United States Government, 1940-1945," Washington, D.C.: U.S. Government Printing Office, August 1945, 40¢.

for the instantaneous diffusate, called the 'ideal separation factor'  $\alpha$ , is given by

$$\alpha = \sqrt{\frac{m_2}{m_1}},$$

where  $m_1$  is the molecular weight of the lighter gas and  $m_2$  that of the heavier. Applying this formula to the case of uranium will illustrate the magnitude of the separation problem. Since uranium itself is not a gas, some gaseous compound of uranium must be used. The only one obviously suitable is uranium hexafluoride,  $\text{UF}_6$ , . . . Since fluorine has only one isotope,\* the two important uranium hexafluorides are  $\text{U}^{235}\text{F}_6$  and  $\text{U}^{238}\text{F}_6$ ; their [approximate] molecular weights are 349 and 352 [respectively]. Thus, if a small fraction of a quantity of uranium hexafluoride is allowed to diffuse through a porous barrier, the diffusate will be enriched in  $\text{U}^{235}\text{F}_6$  by a factor

$$\alpha = \sqrt{\frac{352}{349}} = 1.0043."$$

Note that the enrichment is only 0.43% over the previous proportions.

Some of the difficulties of this process of separation are evident, not the least being the need for the final chemical extraction of uranium from uranium hexafluoride. Another serious point is that this separation factor  $\alpha$  is far too low to permit use of the feebly enriched uranium hexafluoride after one such process, therefore the first diffusate is made to go through another porous barrier which again discriminates against U-238 in favor of the prized U-235. Since  $\alpha$  remains about the same, each single stage changes the relative abundance of U-235 by only very little. In Smyth's words:

"To separate the uranium isotopes, many successive diffusion stages (i.e., a cascade) must be used. . . . Studies by [K.] Cohen and others have shown that the best flow arrangement for the successive stages is that in which half the gas pumped into each stage diffuses through the barrier, the other (impoverished) half being returned to the feed of the next lower stage. . . . If one desires to produce 99 percent pure  $\text{U}^{235}\text{F}_6$ , and if one uses a cascade in which each stage has a reasonable over-all enrichment factor, then it turns out that roughly 4000 stages are required."

The construction by 1945 of many acres of such diffusion chambers in a Tennessee valley (Clinton Engineering Works, Oak Ridge, Tenn.; Fig. 20-4) is very excitingly described in the report. In a

\* Of relative molecular weight 19.



FIG. 20-4. Diffusion plant of the Clinton Works (now Atomic Energy Commission) at Oak Ridge, Tenn., for the separation of uranium isotopes.

certain sense, the first atomic bomb which used U-235 was then a gigantic confirmation of the kinetic theory which produced Eq. (20-13) and this diffusion process.

Yet a third proof of Eq. (20-13), on a rather more modest scale, is the closer observation of Brownian movement. If we watch the thermal agitation of a single tiny dust particle (mass  $m_1$ , rms velocity  $c_1$ ), we note that its erratic dance movement never reaches the excessive speeds which we know the surrounding air molecules possess. In fact, Eq. (20-13) assures us that the rms speed  $c_1$  of the dust particle is equal to  $c_2\sqrt{m_2/m_1}$ ; and since  $m_2$ , the mass of the air molecules, is many million times smaller than  $m_1$ , the mass of our dust particle, we cannot be surprised that  $c_1$  is so small.

Bodies larger than dust particles and the pollen of flowers are, of course, also randomly bombarded by the surrounding atmosphere, but because their masses  $m_1$  are larger still, their  $c_1$  is correspondingly smaller and their Brownian movement becomes unnoticeable. Yet very delicately suspended light mirrors, as used in accurate instruments, do show the slight random jitter of Brownian movement.

(d) Consider another attack on Eq. (20-11),  $T = \frac{2}{3k} \cdot \frac{m_0\bar{v}^2}{2}$ . For

any given gas we can measure  $T$ , we know its mass per molecule  $m_0$ ,\* and we have calculated  $k$  before; thus  $c \equiv \sqrt{v^2} = \sqrt{\frac{3 kT}{m_0}}$ . If we now were able to measure  $\sqrt{v^2}$  directly, we could check Eq. (20-11) at once. But how are we to make a measurement of the incredible speed of invisible molecules?

Exactly because a scientific problem looks most paradoxical and difficult, good minds will ponder it and frequently produce most ingenious solutions. One of the most direct solutions (Otto Stern, 1920, and others) to our problem (in schematic outline) requires that atoms of a metal be evaporated in all directions from a hot wire of known temperature. The wire runs down the center of a close-fitting hollow cylinder which has a thin slit that can be opened and shut like a fast shutter. When the shutter is briefly uncovered, a short, thin, well-defined "squirt" of atoms escapes into the evacuated space beyond the cylinder; as they continue undisturbed in a straight line, the fast atoms outdistance the slower ones, thus spreading into a procession what started out to be a thickly packed crowd of atoms.

If now all atoms are allowed to impinge on a screen or slide of glass, they will quickly condense on it and form a thin coating of metal at a well-defined spot. But in order to separate on the screen the fast atoms from the slow ones, the screen is pulled across the path of the atoms as they fall on it; thus the coating becomes a lengthy smear instead of one spot, with the fastest atoms falling at the head of the deposit, the slowest ones at the tail-end.

This is rather like shooting off a load of buckshot against a moving train at the very moment when the first car passes by. The speediest pellets will indeed hit the first car, and the slowest might hit the last. From the distribution of hits along the side of the whole train, you could analyze the distribution of velocities of all pellets.

One cannot simply and directly count how many atoms fall on any part of the moving plate, but from the geometry and known motion of the slide in this arrangement, we can calculate the absolute speeds (in cm/sec) needed by an atom to reach a given portion of the screen. By mounting the slide and the shutter on the inside

\* The mass  $m_0$  in grams of a gas molecule is given by

$$m_0 = \frac{\text{mass in grams per mole of gas}}{\text{number of molecules per mole}} = \frac{\text{number of grams/mole}}{6.02 \times 10^{23} \text{ molecules/mole}}$$

by Avogadro's law. For nitrogen, for example, this comes to

$$m_0 = \frac{28}{6.02 \times 10^{23}} = 4.65 \times 10^{-23} \text{ gm/molecule.}$$

of two concentric rotating drums with the oven at the center (see Fig. 20-5), one can accurately repeat the procedure every time that these drums make one revolution, and gather many such layers of spread-out "squirts" of atoms on the same plate. Then the relative opaqueness, the measurable *density* of the total smear tells us directly what *fraction* of atoms had what speeds.



FIG. 20-5. Schematic representation of apparatus for measuring the velocity distribution of gas molecules.

The very first thing that a qualitative examination of such a smear reveals is the presence in the stream of atoms of all speeds, from practically zero to practically infinitely large. The percentage of atoms (i.e., the relative density of the smear) at the extreme values is almost negligibly small but rises progressively as we shift our attention to the more central region of the deposit. For silver atoms evaporated from a filament kept at a temperature several hundred degrees below the actual melting point of the metal, the maximum density of the silver coating in the original experiments fell on that portion of the slide which corresponds to a speed of about  $5.5 \times 10^5$  cm/sec. This speed we may call the *most probable speed*, but it is not necessarily the same as the rms speed. To determine the latter, we divide the deposit on the slide into small regions and measure more exactly what fraction or percentage of the atoms had speeds of  $0.1 \times 10^5$  cm/sec,  $1.2 \times 10^5$  cm/sec,  $2.3 \times 10^5$  cm/sec, etc. We might find that for the given vapor and temperature, within the error of measurement, the following *distribution* of velocities exists.\*

Velocity range in $10^5$ cm/sec	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	above 13
% of atoms in range	0.5	3	7	11	15	16.5	14	11.5	9	7	3	2	1	0.5

\* These values actually apply only to truly random speeds such as exist in a gas or vapor in a closed box. The atoms in the *beam* described above have a slightly different distribution because the shutter arrangement automatically gives preference to one type of atoms—those with high velocity components toward the slit. The values in the table are thus really deduced indirectly from the actual density measurements, but this need not concern us here.

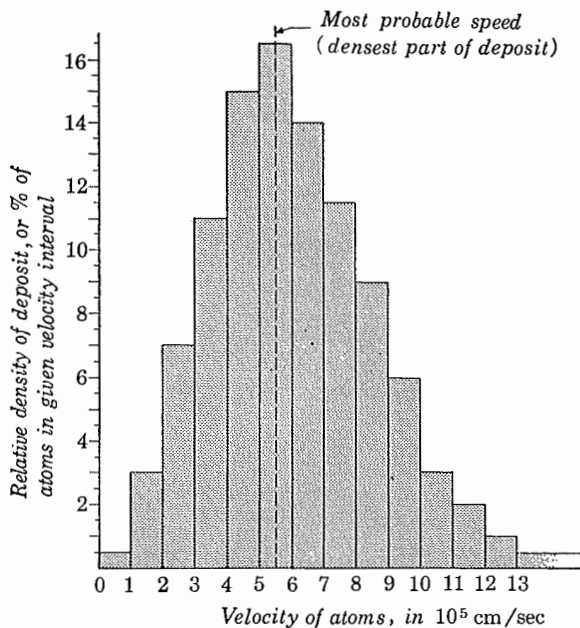
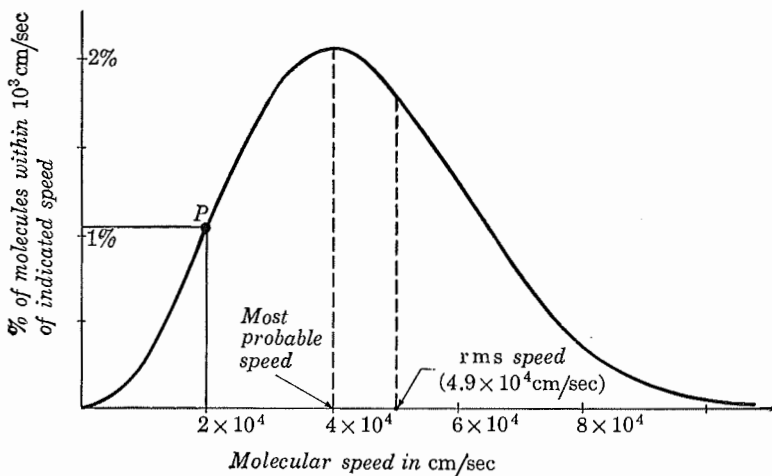


FIG. 20-6. Velocity distribution of atoms in silver vapor.

FIG. 20-7. Distribution of speeds for nitrogen molecules at  $0^\circ\text{C}$ .

This may be represented by a plot, as in Fig. 20-6. We see at once from the figure that (a) the observed speeds are bunched strongly around the maximum, the "most probable" value, dropping off to about 20% of the maximum at twice the most probable speed, and (b) that the distribution of speeds is quite lopsided toward the higher values of speed. Even the straight average speed  $\bar{v}$  is higher than the most probable speed—as you may ascertain by calculation.

PROBLEM 20-5. Is it plausible that the atoms in a chemically homogeneous gas should have different *speeds*? (*Hint*: Consider what would happen if all atoms at one instant *did* have equal speeds; would this situation last if they collided with one another?)

PROBLEM 20-6. The problem of averaging is important in all sciences. As an example, find the (approximate) average speed  $\bar{v}$  from the data pertaining to Fig. 20-6. [*Hints*: To orient yourself for this kind of a problem, consider first how to get an *average* molecular weight for natural pure uranium, which consists of 99.3% U-238 and 0.7% U-235 (except for some negligibly small traces of U-234. Neglect also the fact that U-238 and U-235 actually have relative molecular weights just a little larger than 238 and 235 respectively).]

What our whole experiment tries to prove, however, is that the rms speed,  $\sqrt{v^2}$ , obtained by calculation from these measurements, coincides with the theoretical value predicted by the kinetic theory in the derived equation

$$\sqrt{v^2} = \sqrt{\frac{3 kT}{m_0}}.$$

Here is the final result. If we carry through the calculation of the rms speed based on these measurements (how is this to be done?), our experiment yields a value of about  $6.5 \times 10^5$  cm/sec; on the other hand, the calculated value of  $\sqrt{v^2}$ , using the derived equation for this gas and for this temperature, gives us the *same value* within the expected limits of error!

This indeed is a striking confirmation of our theory. But it is even better than may appear here, because our theory may actually be used to predict the very shape of the distribution curve of speeds—this task was solved through the genius of James Clerk Maxwell in 1859—a prediction which is indeed fulfilled, but which it is far beyond the scope of this treatment to derive. Parenthetically, note also that  $\bar{v}$  has a value very close to  $\sqrt{v^2}$ , as we assumed some time ago.

Before we turn to another achievement of kinetic theory, let us briefly reformulate our results. For the sake of more accurate computations, we may plot Fig. 20-6 again for thinner strips and smaller velocity intervals, until the stepwise graph merges into a smooth line as in Fig. 20-7, drawn here for nitrogen gas at  $0^{\circ}\text{C}$  and for speed intervals of  $10^3$  cm/sec. The form of this interesting curve can be rendered by the equation  $y^2 = (\text{a constant}) \cdot x^2 \cdot e^{-x^2}$ , where  $e = 2.718$ . The meaning of any

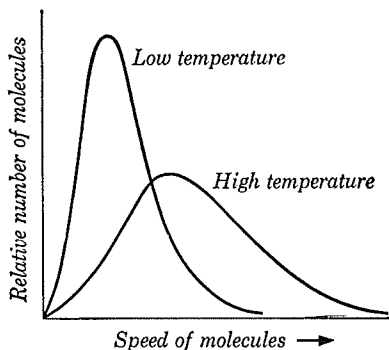


FIG. 20-8. Plot of speed distribution of the molecules of one gas at two different temperatures.

point on the curve, say  $P$ , is now the following: In this gas at this temperature only about 1.1% of all molecules have speeds in a  $10^3$  cm/sec interval centered around  $2 \times 10^4$  cm/sec (i.e., have speeds of  $2 \times 10^4$  cm/sec  $\pm 0.5 \times 10^3$  cm/sec). Figure 20-8 represents the speed distribution in a given gas at two different temperatures.

A useful concept arising from this discussion is *probability*. We may say that the probability is 0.011 that any randomly selected nitrogen molecule at  $0^{\circ}\text{C}$  will happen to have, within  $\pm 0.5 \times 10^3$  cm/sec, a speed of  $2 \times 10^4$  cm/sec (see Fig. 20-7). As we use the words here, "probability of 0.011" means therefore that previous observations of the several speeds of many molecules make it safe to predict that of the next randomly selected 1000 molecules, about 11 will have the specified speed range. If we look at the speed of only 100 molecules, perhaps one will have the desired speed, but the smaller our sample becomes, the less predictable are the results. For a *single* molecule we cannot predict any definite speed, except perhaps to offer the platitude that the probability is 1.00 (100%) that its speed lies somewhere between zero and infinity; in fact, this whole concept of probability does not apply to such a singular situation.

**PROBLEM 20-7.** From Fig. 20-7, find the probabilities corresponding within the given range to the most probable speed, to the rms speeds, and to twice and three times the most probable speeds.

Note that although the probabilities are smaller and smaller at extreme (high and low) speeds, they are not totally zero; that is, a



few precocious molecules could be found at all times (and temperatures above absolute zero) which will show enormous speeds and consequently enormous energies, whereas at any moment a few others will seem practically to stand still. Similar distribution curves apply to liquids and even to solids, and they explain why some molecules, even at low temperatures, will be able to escape ("evaporate") from the surface.

(e) An important implication of Eq. (20-11) is that the  $\overline{KE}_{\text{trans}}$  of a perfect gas decreases linearly as the temperature (in °K) is lowered toward zero; this gives physical significance to "absolute zero temperature" as the temperature where—to put a subtle point very crudely—the random thermal motion of gas molecules has ceased.

(f) Again on the qualitative side, one of the very important purposes of the kinetic theory was to explain why a gas in a thermally insulated container (adiabatic process) rises in (absolute) temperature in proportion to the work done on it by reducing its volume, e.g., by means of a piston pushing in one side of the container. In fact, it was this type of observation which first served as a foundation for the comprehensive statement of the law of conservation of energy. The work done on the gas, or the energy given to it by the advancing piston, is equal to the increase of translational kinetic energy picked up by the gas molecules during collision with the moving piston. But by Eq. (20-11), any change in  $\overline{KE}_{\text{trans}}$  is reflected in a proportional change in the temperature:  $\Delta\overline{KE}_{\text{trans}} \propto \Delta T$ ; hence the proportional rise in temperature.

The same argument in reverse explains why an (idealized) gas cools itself when it is allowed to expand—not into a vacuum, in which case it does no work and, by experiment, suffers no total temperature drop, but when it expands against a receding piston, or the inside walls of a balloon, or against another gas at lower pressure, or some "obstacle" of that sort. A familiar example is presented by a warm mass of air rising to a higher region, expanding against the more rarefied atmosphere, and so cooling itself to a temperature where the moisture brought along in the air is forced to condense into a rain or snow cloud.

Turning briefly again to the kinetic theory as applied to liquids, we now realize why evaporation from the surface cools the body of the liquid: since only the fastest molecules overcome the surface attraction and escape by evaporation, the average KE of the remaining ones drops correspondingly; however, a drop in  $\overline{KE}_{\text{trans}}$  is, in fact, equivalent to a drop in  $T$ .

PROBLEM 20-8. Equation (20-11) can be written as  $\frac{3}{2} kT = \frac{1}{2} m_0 \bar{v}^2$ . Thus the  $\overline{\text{KE}}_{\text{trans}}$  per molecule of any gas is  $\frac{3}{2} kT$  and, for any gas, the

$$\text{total KE}_{\text{trans}} \text{ per mole} = \left(\frac{3}{2} kT\right)N = \frac{3}{2} RT \quad (20-14)$$

(a) Calculate this energy (in ergs) present in one mole of water vapor (superheated steam) at 200°C; also find the amount of energy given up (to the receding piston) by 9 gm of expanding steam in a steam engine cylinder, if the steam cools itself during an adiabatic expansion to 100°C (all heat losses to the surroundings neglected). (b) Calculate also the amount of energy that must be removed and might be usefully available if 1 mole of ordinary air were cooled from room temperature to about -200°C, where these gases liquefy. Why don't we ordinarily utilize this energy to do useful work?

(g) An important result of the kinetic theory, on which we can only touch briefly, was that it permitted a fair estimate of the *size* of individual molecules. Recall that Dalton tried to estimate the relative sizes of atoms on the basis of the mistaken model of a static gas in which the atoms touched. In the kinetic theory we have been assuming throughout that the volume of the molecules is negligible, seemingly destroying all hope of estimating sizes. However, there are several phenomena which can be treated by our model without making that assumption, specifically, (1) the pressure-volume behavior of denser gases, and vapors, where the molecular volume becomes of importance and where consequently Boyle's law breaks down; (2) the speed with which one gas diffuses through another, where evidently the size of the atoms will determine how often they collide and how easily and quickly they will pass by one another; (3) the speed with which heat can be conducted through a gas, another instance where the size of atoms will determine their progress despite collisions; and (4) the measurable viscous drag which a gas exerts on a larger object moving through it, and which again brings the factor of molecular size into the calculations. Suffice it to say that each of these problems involved enormous theoretical difficulties, that some are still the subject of fundamental investigation, and that all lines of evidence, from the 1860's on, converged on the result that the "diameters" of simple gas molecules are of the order of  $10^{-8}$  to  $10^{-7}$  cm—a conclusion in harmony with our assumption in Section 20-3.\*

\* Evidently we must be on guard against conceiving of too rigorous a model for molecules. The "diameter" may refer only to some average distance of closest approach among molecules. For a qualitative account of the determinations of size and mean free path of molecules, see T. W. Chalmers *Historic Researches* London: Morgan Brothers, Ltd. (1949) page

(h) Because it was Dalton's question of why the atmosphere's constituent gases are so thoroughly mixed which started him on his road to the atomic theory, we should surely give a brief look at the solution now given by kinetic theory to this problem. You recall that Dalton believed the mixing to be a result of a brief initial period of disequilibrium among unequally large, mutually repelling atoms, and that he was wrong on several counts. The kinetic theory picture of gases predicts that two or more gases, intermixed like the atmosphere, should diffuse one through the other by virtue of the thermal movement of the molecules. However, and quite ironically, the heavier gas (namely, oxygen) is indeed expected to be found relatively somewhat more frequently per 100 molecules at lower levels than at levels higher in the stratosphere (why?). While this does not correspond at all to the marked stratification between a heavy and a light liquid—say water and oil—it is nevertheless just the type of phenomenon which Dalton thought does not happen. And there is a further twist: whereas the expectations of the kinetic theory on that subject are fulfilled in a suitable quantity of gas under ideal conditions, our own atmosphere is so disturbed by wind currents that there is a fairly even mixing of all gases throughout the region accessible to direct and indirect measurements; so that it is after all a disequilibrium in the gas which explains the observations, but a continuous one, of a type not looked for in the Daltonian context.

(i) A startling set of observations to which the kinetic theory lends direct explanation concerns the *escape of the atmosphere* from planets and satellites. As everyone knows, the moon and the smaller planets, with their smaller gravitational pull, have long ago lost their atmospheres—if they ever had any—and our earth slowly is doing the same, although we may be sweeping up enough interplanetary gas and vapor from meteorites to balance the loss. The process of escape is evident: The finite weight of the atmosphere "packs" it more densely near the earth, but the density decreases rapidly with height. The distance traveled by each molecule between collisions becomes correspondingly greater—at sea level it is only  $10^{-5}$  cm, but at a height of 1000 km it may be many miles. The paths of molecules will then be ordinary trajectories, but if their "initial speeds" are large enough (over  $10^6$  cm/sec at that height above our earth) they may shoot too far out to be pulled back by the diminishing gravitational field. One glance at Fig. 20-7 shows that for nitrogen at  $0^\circ\text{C}$  there are only very few molecules indeed having

---

156 ft. A fine technical account is in T. G. Cowling's *Molecules in Motion*.

such speeds. However, there are always some molecules at all speeds, and the temperatures at that height being much greater (perhaps 1000°C) the rate of escape for the common atmosphere is a little larger than would first appear, although still practically negligible. Finally, the smaller the molecular weight the larger the proportion of high-speed molecules at any given temperature, which means, in the extreme, that a hydrogen molecule is likely to escape from our earth only about 1000 years after it is liberated (in chemical decomposition).

(j) We now come to one of the most striking and intriguing applications of the kinetic theory of gases, the explanation of the specific heats of gases.

To find the specific heat of a gas (see Section 17-8), we may take one gram of the gas, put it in a rigid, heat-insulated container, and feed it a measured amount of heat energy (in calories) until its temperature rises by one degree (on the centigrade *or* the absolute scale—in both, one degree corresponds to the same *interval*). The specific heat\* of the gas  $s_v$  = calories needed per gm of gas, per °C of temperature rise. This experiment would yield widely different values for different gases, 0.75 (cal/gm)/degree for helium, 2.41 for hydrogen, 0.15 for carbon dioxide, etc. Instead of using one gram, we might have used one mole of the gas, defining thereby another but related concept, the *molar* specific heat,  $C_v$ , i.e., the number of (calories/mole)/degree.  $C_v$  would correspondingly differ from the values of  $s_v$ ; evidently  $C_v = s_v \times$  number of grams per mole of gas =  $s_v \times$  relative molecular weight of the gas.

In either case, our model of the gas tells us that for every calorie we supply the total  $KE_{\text{trans}}$  of the molecules will rise by  $4.18 \times 10^7$  ergs (law of conservation of energy); for  $Q$  calories of heat energy supplied,  $KE_{\text{trans}}$  increases by  $\Delta KE_{\text{trans}} = QJ$  ergs (where  $J$  is  $4.18 \times 10^7$  ergs/calorie, the mechanical equivalent of heat).

Consider 1 mole of gas, about to receive  $Q$  calories. By Eq. (20-14), the total  $KE_{\text{trans}}$  per mole of this gas =  $\frac{3}{2} RT$ . Since the temperature and the total  $KE$  are linearly proportional,  $\Delta KE_{\text{trans}} = \frac{3}{2} R \Delta T$ ; let us write this  $\Delta KE_{\text{trans}}/\Delta T = \frac{3}{2} R$ . From the law of conservation of energy,  $\Delta KE_{\text{trans}} = Q \cdot J$ ; thus  $QJ/\Delta T = \frac{3}{2} R$ , and therefore we write

$$\frac{Q}{\Delta T} = 3 \frac{R}{2J}. \quad (20-15)$$

\* At constant volume, care being taken that none of the heat energy supplied is wasted in work done by an expansion of the gas.

The left-hand side is the heat energy (cal) supplied to this one mole per degree rise of temperature—but that is exactly the definition of  $C_v$  for a gas.

$$C_v = 3 \frac{R}{2J}. \quad (20-16)$$

We note with astonishment that by our theory the molar specific heat of all gases should be the same,  $R$  and  $J$  being universal constants! Specifically, substituting values for  $R$  and  $J$ , we get

$$C_v = 3 \times \frac{8.31 \times 10^7}{2 \times 4.18 \times 10^7} = 3 \text{ (cal/mole)/degree (very nearly).}$$

We immediately rush to determine  $C_v$  experimentally—and indeed, for helium, argon, and mercury vapor,  $C_v$  is 3.02, 2.95, and 2.94 respectively, and the values for the other monatomic gases also confirm our theory within tolerable limits.

But only the *monatomic* gases! Diatomic ones such as hydrogen, nitrogen, oxygen, carbon monoxide, and hydrogen chloride have experimental values of  $C_v$  near 5 (cal/mole)/degree instead of the predicted value of 3. For these it is as though our theory must somehow be modified to yield

$$C_v \text{ (diatomic gases)} = 5 \frac{R}{2J}.$$

In short, diatomic molecules seem to have a way of absorbing (receiving and storing) more energy per degree temperature rise than do monatomic molecules. Every calorie supplied to molecules made of only a single atom each did turn up as translational energy, as expected; but every calorie supplied to a diatomic molecule seems to split 5 ways—only 3 parts go into translational energy and occasion the observed temperature rise, and the remaining 2 parts somehow are stored away without affecting the  $KE_{\text{trans}}$  or the temperature. What is the physical explanation of this?

This is simply a breakdown of our model. Joule, who brought his work to this point, was stumped by the discrepancy in the specific heats of polyatomic gases, the more so as in his day no monatomic gases had yet been generally identified. In fact, his initial thought was that the experimental values of  $C_v$  then available might be in error; and when new determinations only widened the gap between prediction and experimental results, he seems to have been so discouraged as to leave the whole subject of kinetic theory. He had not made any provision for a molecule to receive other energy than translational kinetic energy, and up till now there was no reason to

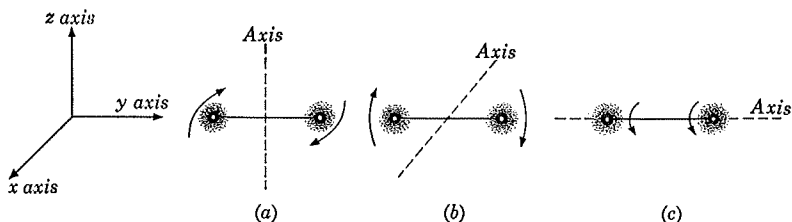


FIG. 20-9. Three independent ways in which a diatomic molecule can rotate.

revise this simple idea. Now, however, we search for a plausible mechanism to account for the unsuspected heat capacity of diatomic molecules. And it might occur to us (this was done first by Clausius in 1857) that even if monatomic gases seem to fill the role of agitated point masses or small spheres without *rotational* energy, diatomic molecules might well be thought of as dumbbell-like objects, as two small spherical masses a short distance apart, an arrangement which has a much larger rotational inertia than either of the individual "point masses," and which therefore is much more capable of being put into rotation and of containing a significant amount of rotational kinetic energy ( $\frac{1}{2} I\omega^2$ ) at plausible speeds of rotation.

Figure 20-9 shows a very crude schematic presentation of a diatomic molecule and its three possible types of rotation, distinguished from one another by the positions of the axis of rotation—analogueous to our method of differentiating between the three possible types of translational motions, i.e., motion along the *x*, *y*, or *z* axis.

**PROBLEM 20-9.** Compare the rotational inertia of a hydrogen molecule ( $H_2$ ) conceived (a) as a small sphere, as in our first model, and (b) as a dumbbell molecule with two equal (spherical) hydrogen atoms at about one atomic diameter apart. [Note:  $I$  depends on the axis of rotation (see Fig. 20-9). A dumbbell molecule has a different value in (a) than in (c).] Numerical values for  $I$  are obtainable if we make the assumption, based on other experiments, that the diameter of a single H atom is *about*  $10^{-8}$  cm, and that all but about 1/2000 part of the total mass of the atom is concentrated at its center, the nucleus, whose radius is perhaps a  $10^{-4}$  part of the atom's total radius.

It is very convenient to invent the term "*degree of freedom of motion*" to denote each one of the independent ways in which a molecule can receive energy. For example, a monatomic gas molecule which does *not* rotate, as we initially pictured it, can acquire kinetic energy only by virtue of its velocity along the *x* axis or *y* axis or *z* axis. These three motions are independent components, in the

sense that we could imagine that one component is increased or decreased without the others being affected. In other words, the molecule has three degrees of freedom. If, however, we consider a particle made of two or more atoms, capable of rotational as well as translational motion, it has three degrees of freedom for translation *plus* three more for the three possible independent ways of rotating (about the three different axes shown in Fig. 20-9), or six degrees of freedom altogether.

However, we must be prepared to face the possibility that a molecule for some reason may not make use of all the conceivable degrees of freedom to which it is entitled by simple geometry. Thus we have already implicitly denied a monatomic molecule the use of its own three geometrically quite conceivable degrees of freedom of rotation or spin by saying that its rotational inertia is too small to make them effective, or that it is a point mass—a rather dubious trick, to be re-examined later, and, in fact, a veiled way of saying that in order to explain the experimental facts of specific heat in terms of our model we need not and must not allow rotation for monatomic gas molecules in our model. Similarly, if rotation is neglected for a single atom, it should also be neglected for that one degree of freedom of rotation of a diatomic molecule which corresponds to motion about the lengthwise axis [Fig. 20-9(c)]; therefore molecules of  $H_2$ ,  $N_2$ ,  $O_2$ ,  $CO$ ,  $HCl$ , etc. should have only five degrees of freedom (three for translation, two for rotation), or two degrees more than monatomic molecules.\*

Here, then, we have an explanation why a mole of diatomic gas absorbs more energy per degree of temperature rise than does a mole of monatomic gas. Of any and every unit of heat energy supplied to the former, only  $\frac{3}{5}$  goes into translational motion [ $\Delta KE_{\text{trans}}$ , which by Eq. (20-11) causes a proportional rise in temperature], and the remaining  $\frac{2}{5}$  of the supplied energy is stored in the two active modes of rotation ( $\Delta KE_{\text{rotation}}$ , which does not directly influence the rise in temperature). Therefore, to raise the temperature of a mole of diatomic gas by one degree, we must give it as much energy as that required for one mole of monatomic gas *plus* almost as much again ( $\frac{2}{5}$  of the total) to satisfy the two active rotational degrees of freedom. So we understand that  $C_v$  (monatomic gas) =  $3 (R/2 J)$ , whereas  $C_v$  (diatomic gas) =  $5 (R/2 J)$ .

\* The actual existence of rotational motion of polyatomic molecules, which affects neither the pressure nor the temperature, has been confirmed independently by direct experimentation on Brownian movement and on the light emitted by rotating molecules.

Reinspection of these two equations now reveals some interesting facts: (1) If we were dealing with polyatomic gases whose molecules are so complex as to have significant rotational inertias about all three axes of symmetry, we should expect the third degree of freedom of rotation also to play its part. With a total of six active degrees of freedom, we expect  $C_v$  (polyatomic) =  $2 \times C_v$  (monatomic) =  $6 (R/2J)$ , or about 6 (cal/mole)/degree. In actual fact, many triatomic gases *do* have values of  $C_v$  of about 6 (5.9 for  $\text{CO}_2$  at low temperatures, 6.1 for water vapor at  $100^\circ\text{C}$ ). Other polyatomic gases have even higher values, indicating that further hitherto neglected degrees of freedom begin to be important, namely, those representing *vibration* between the components of the molecule. For example, the complex molecules of ethyl ether ( $\text{C}_4\text{H}_{10}\text{O}$ ) have a molar specific heat  $C_v$  of about 30. Thus the idea of degrees of freedom of motion superposed on our initial model of the gas allows us to interpret the experimental facts of specific heat very satisfactorily. Within reasonable limits, we may write

$$C_v = \delta \left( \frac{R}{2J} \right), \quad (20-17)$$

where  $\delta$  = the number of active degrees of freedom of motion for the specified gas molecules.

This notion can be expanded to explain the observed specific heats of solids in terms of kinetic theory. But note now the astonishing fact that the specific heat of a gas has just given us a strong clue as to the molecular structure of gases, the very information Avogadro had lacked to protect his theory from the worst doubts. For example, the fact that  $C_v$  for hydrogen gas is ordinarily 5(cal/mole)/degree shows that it must be diatomic—not monatomic as Dalton had thought. Similarly,  $C_v$  for mercury vapor is about 3 (cal/mole)/degree—clear evidence for its monatomic structure even though the other (chemical) evidences may be ambiguous or difficult to obtain.

(2) Implied in the last few paragraphs was an important idea, the *equipartition of energy* among all active degrees of freedom in a gas: we have assumed that any and every unit of heat energy supplied to a mole of gas is distributed in equally large shares to each active degree of freedom. In terms of individual molecules this means that at a given temperature each gas molecule has, on the average, energy in amount equal to  $\frac{1}{2} kT$  for each of its active degrees of freedom (about  $1.88 \times 10^{-14}$  erg at  $0^\circ\text{C}$ ). Hence a typical average



monatomic molecule (with  $\delta = 3$ ) possesses a total energy of  $\frac{3}{2} kT$  per molecule (translational energy only, and thus equal to  $\frac{1}{2} m\bar{v}^2$ ), diatomic ones each have  $\frac{5}{2} kT$ , polyatomic ones have even more. If any external energy  $\Delta E$  is supplied to the gas, each active degree of freedom absorbs an equal share of  $\Delta E$ . This principle of equipartition, which we have introduced to explain the discrepancy in specific heats, is derivable more rigorously from the experimental facts, and is of foremost importance in the study of gases.

(k) Finally, we should examine the points where our kinetic theory frankly breaks down. We expected such a possibility from the beginning, because our gas model was admittedly very simplified, and the introduction of the idea of degrees of freedom and of equipartition of energy may not have had enough sophistication in the long run. And not only did we expect the inevitable deficiency which appears sooner or later in every theory, but we eagerly look for it as a clue to the necessary modifications and perhaps as a key to the strange laws governing the world of atoms. One-third of the excitement in the study of physics lies in constructing a theory, another third in confirming the theory, and yet another in pushing beyond the theory.

The outstanding defect in our theory can be summarized in the following paradox: the principle of equipartition of energy requires that equal amounts of energy exist at a given temperature in each degree of freedom—except in those (the inactive) degrees of freedom where *no* energy at all resides, e.g., in the geometrically possible rotational degree for diatomic molecules about the lengthwise axis. We are forced to ask, *exactly what determines whether a degree of freedom should be active or inactive?*

This question becomes even more intriguing if we note a discovery (made in 1912) concerning the amazing behavior of hydrogen gas,  $H_2$ . At ordinary temperatures its value of  $C_v$  is a little below 5 (cal/mole)/degree, as becomes a diatomic gas. But at lower temperatures,  $C_v$  for this gas falls to about 3 (cal/mole)/degree (at about  $100^\circ K$ ), the value appropriate for a monatomic gas (see Fig. 20-10)! We must conclude that the two rotational degrees of freedom have "frozen in." Nothing in our theory has prepared us for such shocking behavior.\* And what is worse, on heating  $H_2$  gas or others such as  $O_2$  to extremely high temperatures,  $C_v$  increases to over

\* If, on cooling, the two nuclei of  $H_2$  came very close together, we might perhaps accept this as an explanation; but the molecular dimensions change only very little during cooling, as we find from other lines of investigation.

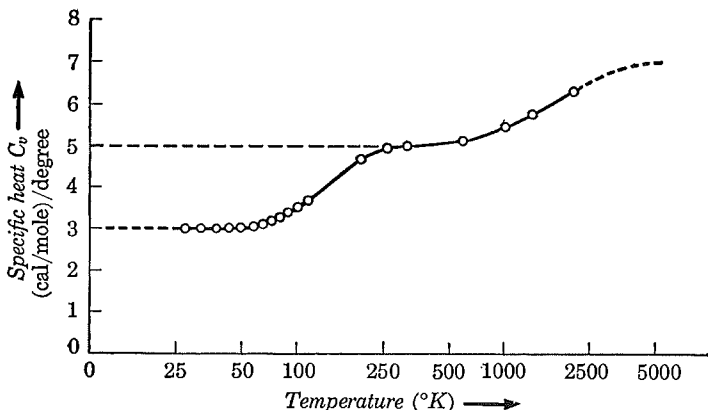


FIG. 20-10. Variation of specific heat  $C_v$  of molecular hydrogen gas. (To condense the graph, the temperature of the gas is plotted along the abscissa on a logarithmic scale.)

5 (cal/mole)/degree (at about 2000°C), as though other degrees of freedom begin to play a part not usual for diatomic molecules at ordinary temperatures.

The fact of the matter is that here we come into entirely new and unexpected territory; this will be our link between the “classical” concept of the atom as formed in kinetic theory and the “modern” concept which the next part of the text will develop. But we may indeed rejoice how well our simple mechanical model now seems to serve us in so many contexts (with that exception of specific heats). So successful was the “billiard-ball” picture of the atom that it seemed that any revision should prove only minor. That hope, however, was in vain—the revision was accompanied by a fundamental revolution of physics itself. This is the topic to which we now turn.

### Additional Problems

PROBLEM 20-10. If  $H_2$  gas above 2000°C had, on the average, 3 active degrees of freedom of translation plus 2 of rotation and 2 of vibration, what is the  $\overline{KE}_{\text{trans}}$  energy of each molecule, and what is the value of  $C_v$ ?

PROBLEM 20-11. In the light of the previous problem, what could be the physical explanation of the observed value  $C_v = 7.6$  (cal/mole)/degree for gaseous sulfur dioxide ( $SO_2$ ) at ordinary temperatures?

PROBLEM 20-12. Review the discussion leading to Eq. (20-8) and point out the various laws of physics openly or tacitly employed in the derivation.

PROBLEM 20-13. Assume that you have discovered a new gas that on being supplied with heat (from a Bunsen burner) raises neither the pressure

in the vessel nor the temperature of the gas; what can you say about motion of the molecules of this gas?

### Further References

T. W. Chalmers, *op. cit.* (Ch. 17). Chapter VIII.

T. G. Cowling, *Molecules in Motion*. London: Hutchinson's University Press (1950).

H. Crew, *op. cit.* (Ch. 2), Chapter IX.

L. B. Loeb and A. S. Adams, *The Development of Physical Thought*. New York: Wiley (1933). Part III, an excellent discussion of "Heat and the Structure of Matter."

W. F. Magie, *op. cit.* (Ch. 2), pp. 165-174, 251-261. Excerpts from Gay-Lussac, Joule, Thomson, Brown, and Maxwell.

*Part G*

QUANTUM THEORY  
AND THE NUCLEAR ATOM

- 21 Electrostatics
- 22 Properties of Light
- 23 The Quantum Theory of Light
- 24 Radioactivity and the Nuclear Atom
- 25 Bohr's Model of the Atom



*Niels Bohr*

## QUANTUM THEORY AND THE NUCLEAR ATOM

To most scientists of the late 19th Century, the physical world, despite all its superficial complexity, appeared to be endowed with a truly superb unity. The mechanics of Galileo and Newton explained the motion of all bodies—whether everyday objects or planets. Furthermore, the kinetic-molecular theory had triumphantly extended the reign of this mechanics into submicroscopic worlds by interpreting most phenomena of heat as the motion of atoms. Sound, too, had become a branch of mechanics, having been analyzed as a vibrating disturbance traveling among the molecules, according to those same classical laws of motion. Earlier there had been hope that even light might be dealt with as a mechanical wave motion traveling through a material fluid, but although this attractive picture had to be abandoned, its successor did not seem to be in contradiction with classical physics.

There were still many important unsolved problems, but in the 1880's and 1890's one could be confident that eventually the old mechanical concepts would provide the required solution. However, in 1900 there appeared a startling publication by Max Planck that threw doubt upon the most basic concepts. Other key contributions followed within the next few years, from the pens of Einstein, Rutherford, and Bohr. These four men were the standard bearers of a revolution that transformed physics fundamentally. What they did was to change completely the picture of the atom. The revised model gave us the new, larger unity, the new meeting ground of modern science. It not only resolved at one stroke some great problems and many minor puzzles of long standing, but, by providing a common mechanism basic to each of the physical sciences, it also brought together the work of such divergent groups as astrophysicists and industrial chemists. The growth of science became virtually explosive, the harvest in pure and applied knowledge unparalleled, the promise for future developments without limits.

How had this happened? Let us devote this section to that revolution. If our primary aim is again to study the facts and theories of physical science, we shall also find time here for some related tasks. It will become clear, for example, how the new concept of the atom differs from the old, what features in the older theories had all along hidden their own seeds of destruction, and what the experimental facts were which demanded a change in the model as well as in the outlook and methods in the physical sciences.

## CHAPTER 21

### ELECTROSTATICS

**21-1 Introduction.** The physical atom of the 19th century was discussed largely in terms of mechanical force, momentum, kinetic energy, elastic collision, and the like. But as the century drew to a close, there was discovered the electron, that carrier of a fixed negative charge, endowed with a definite mass, a particle much smaller than the atom and undeniably a part of its structure. Then radioactivity brought out more charged particles that had to be somehow accommodated in the atom, re-enforcing long-standing demands from other branches of physics that electric phenomena must be associated with atoms directly. Finally, the 20th century gave us a model that stresses concepts like charge, electric potential, and electric fields at least as prominently as its predecessor had stressed mechanical concepts. Therefore it is not only fitting but necessary to undertake first of all a rather detailed study of the basic ideas of electricity as we now turn from the classical to the modern atomic theory.

**21-2 Electrification by friction.** The word *electricity* is derived from the Greek word for amber, *electron* ( $\eta\lambdaεκτρον$ ), for it was in ancient Greece that the observation seems to have been first made that after amber has been rubbed with cloth it *attracts* small pieces of lint, as we now know, by electrostatic forces. Not until the seventeenth century were forces of repulsion observed in such simple experiments as with bodies electrically charged by rubbing. We still use Franklin's terminology to describe the main observations:

(a) When two different bodies such as a glass rod and a silk cloth are rubbed against each other, they will thereafter attract each other with a new, nongravitational force, because they have developed electric charges of opposite sign. By arbitrary convention we call the charges on the glass positive (+), those on the silk negative (-). Note well that the observed attraction defines two things: that charges are present, and that they are of different kinds.

(b) Two glass rods, rubbed in turn with a silk cloth, both become charged, but repel each other. We note that positive charges repel one another.

(c) When a rod of resin (or amber, rubber, etc.) is rubbed against a new silk cloth, it also becomes charged, as seen by the appearance of a force of attraction between the resin and silk. But this charged resin rod also attracts the previously charged glass rod, wherefore we decide to assign a minus sign to charges on the resin.

(d) Finally, two resin rods charged in the same way repel, and we conclude that negative charges repel one another.

Further experiments with similar combinations reveal no case where the cloth is repelled by the material which it has freshly charged, nor any cases where a charged body is attracted to (or repelled from) both  $+$  and  $-$  charges. All these observations can therefore be summarized by the following hypotheses (and the deduction from the above experiments is a useful exercise in clear thinking):

(i) Two neutral bodies can charge each other by friction, but always with opposite charges.

(ii) There exist only two kinds of charges which interact, *positive* and *negative*. Positive charges appear on glass when rubbed with silk (definition of  $+$  charge), negative ones appear on resin when rubbed with silk or, more effectively, with fur (definition of  $-$  charge).

If suitable care is taken, any object can be charged by friction with a suitable cloth, and the charges thereby produced turn out to be  $+$  or  $-$ , but never of a third kind.

**PROBLEM 21-1.** What is the action of a silk cloth which has charged a glass rod, on another silk cloth which charged a rubber rod? Can you predict what will happen if you rub silk against silk? Silk against fur?

**PROBLEM 21-2.** Although they are not found in nature, some kinds of charges other than  $+$  and  $-$  are logically possible. Describe what properties they would have, i.e., how you would *know* when charges different from both  $+$  and  $-$  are present.

**21-3 Law of conservation of charge.** Even before we define charges quantitatively, we may suspect a law, to be confirmed later more rigorously, of *conservation of charge*, i.e., a law stating that electric charges can be produced only in pairs of equally strong positive and negative charges. The preliminary experimental proof, due also to Franklin, is a simple observation: Wrap an uncharged cloth around an uncharged glass rod. Rub the rod and separate cloth and rod; individually each exerts, as expected, strong electric forces of repulsion or attraction on some third object, perhaps a charged leaf of tinfoil. But now wrap the charged cloth again around

the charged rod; the electric forces exerted on some third object disappear. Our conclusion is that the effect of the opposite charges cancels when they are brought near each other, therefore cloth and rod charged each other equally strongly though with opposite signs. No *net* charge (excess of + or of -) was created by mutual friction. As usual, such a law of conservation is the deepest justification for the power of the concept "electric charge" itself.

**21-4 Modern model for electrification.** Simple experiments like these led to some reasonable and fruitful early theories concerning the nature of electric charges. Our modern model cannot be presented as a logical development without reference to many more subtle and quantitative recent experiments, but let us simplify and speed up our discussion of electrostatics by stating the modern concepts now and justifying them later.

All materials, solid, liquid or gaseous, are atomic in structure, as we convinced ourselves some time ago in discussing the successes of the kinetic-molecular theory. Each atom can ordinarily be considered as a neutral, uncharged object, but actually it is made of a strongly positively charged nucleus (the small central kernel of the atom which represents most of its mass within a radius of about  $10^{-12}$  cm), surrounded at a relatively large distance (about  $10^{-8}$  cm) by negatively charged electrons (from one to over 90 around each nucleus, depending on the particular element). Electrons are all alike, and very light compared with the nuclei.

The neutrality of the ordinary atom is the consequence of an exact balance between the positive charges of its nucleus and the negative charges of its extranuclear electrons. When an electron is somehow torn from its parent atom, the remaining atom lacks one negative electron-charge, i.e., it is a *positive ion*. The separated electron meanwhile represents a free negative charge; it may in time attach itself to a positive ion and so restore the neutrality of that atom, or it may attach itself to a neutral atom and so transform it into a *negative ion*.

In solids, the atoms at the surface are at rest except for thermal vibrations. Their electrons are held near the nuclei by electrostatic attraction, but in some substances it is possible to interfere with this attraction (even by simple surface contact with another material) and to snatch some surface electrons, leaving a positively charged surface on one material and carrying away an equally strong negative charge on the other. This is our picture for electrification by inti-



mate contact, as by friction, although we are not prepared to say just how or why some materials, like glass, joyfully allow their electrons to be carried off, whereas others, like amber and resin, tend to grab extra electrons very copiously.

Here then is a simple model to explain the experimental facts of electrification by friction. Furthermore, it accounts for the observation that there are only two kinds of charge, and it will offer an adequate explanation of all other observed phenomena in electrostatics, e.g., the fact that in principle a practically limitless amount of electric charges can be separated by friction, understandable now in terms of the huge quantity of atoms available on every small portion of surface.

Note that the charges of ions and electrons are not something *apart from* those bodies. For example, it is not possible to discharge an electron and leave a small neutral mass, the whole electron *itself* is a carrier of negative charge. It is best to realize right here that there is no easy picture for visualizing the concept "charge." An electron has a definite mass (this statement undoubtedly and unfortunately invokes a picture of a tiny, hard, round sphere) and the electron also carries a constant negative charge (this statement adds nothing to our picture except that we must endow that little sphere with the ability to repel other electrons and to be attracted by positive ions). In fact, the concept "electric charge" has no meaning apart from some of these nongravitational forces which bodies sometimes exert on one another. The following quotation illustrates the conceptual difficulties in electricity very well:

"Some readers may expect me at this stage to tell them what electricity 'really is.' The fact is that I have already said what it is. It is not a thing like St. Paul's Cathedral; *it is a way in which things behave*. When we have told how things behave when they are electrified, and under what circumstances they are electrified, we have told all there is to tell. When I say that an electron has a certain amount of negative electricity, I mean merely that it behaves in a certain way. Electricity is not like red paint, a substance which can be put on to the electron and taken off again, it is merely a convenient name for certain physical laws." (Bertrand Russell, *ABC of Atoms*, E. P. Dutton, 1923.)

**21-5 Insulators and conductors.** Experimentally we can distinguish between two types of materials: (a) those that retain an electric charge wherever on the surface it was placed initially, and

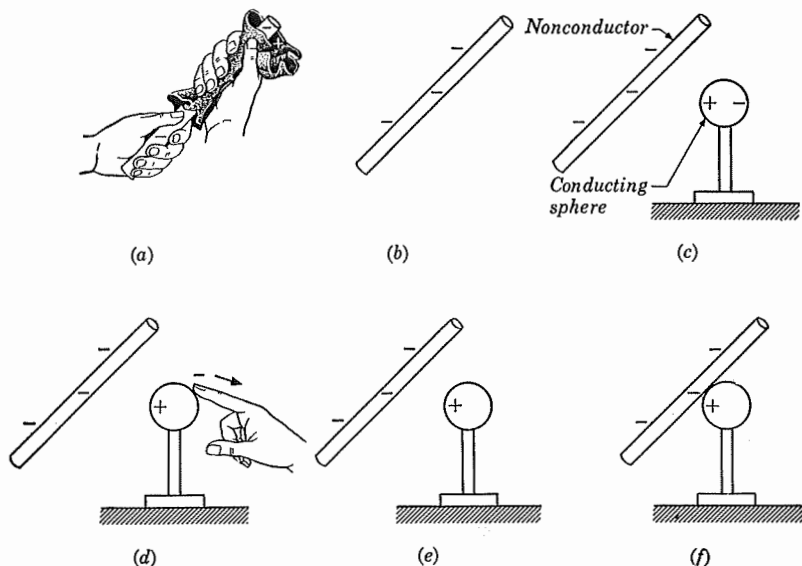


FIG. 21-1.

(b) those that allow equal charges placed on their surfaces to repel one another, to "flow" and distribute those charges all over the body no matter where they were placed initially. The first behavior characterizes *insulators* (or nonconductors, dielectrics) such as glass, fur, silk, paper, resin, etc. The second type of materials, of course, are *conductors* such as metals and graphite.

We can also differentiate between conductors and insulators by the relative resistance which they offer to an electric current. In good conductors like metals some of the electrons more loosely connected to their atoms will quite easily drift away from their parent atoms when some external electric force acts on them; hence they are called "conduction electrons," and while they move they constitute an *electron current* through the conductor (or along its surface). In a good dielectric, however, the electrons are held more tightly by the parent atoms and are not free to flow away when acted on by external electric forces of attraction or repulsion.

Many of the experiments leading to these conclusions can be presented in concise diagrammatic form. Let us agree to indicate by a + or - the sign of the excess of + or - charges at any point on a solid. Then Fig. 21-1 tells the following simple story:

(a) By vigorous rubbing, a piece of fur loses electrons to a resin

or hard-rubber rod; the fur is charged positively, the rod negatively.

(b) The fur is withdrawn, the rod remains charged.

(c) When a conducting metal sphere on an insulating stand is brought near, the conduction electrons are repelled to one side, leaving the sphere differently charged at the two sides; this separation of charges on a body by external electric forces is called *induction*. The rod and sphere will now attract each other, because the rod is nearer to the unlike charges on the left part of the sphere than to the like charges at the right.

(d) We touch the right side of the metal sphere with our finger or with a wire connected to a large conducting body such as the (moist) earth ("grounding"). At once almost all the "free" negative charges at the right flow off, still repelled by the charged rod and by one another, and eager to distribute themselves as far away from the rod and from one another as possible. The positive charges at the left are "bound" charges, that is, they cannot be neutralized by grounding at that side as long as the charged rod holds them in spell. (Explain why.)

(e) Just withdrawing the finger or ground connections at the right leaves that side still neutral, the left still positively charged. If the rubber rod were now also withdrawn, the  $+$  charges, now free, would freely distribute themselves all over the whole metal sphere which is now "*charged by induction.*"

(f) Touching the rod against the sphere at one point also changes practically nothing: none of the excess electrons on the nonconducting rod except those by chance on the point of contact can flow to annihilate the excess  $+$  charges on the sphere. Only if we twist and turn the rod so that in time most of the rod has made contact with the sphere, will each  $+$  charge on the sphere be able to pick up its neutralizing electron from the rod.

**PROBLEM 21-3.** What would have happened if step (c) had been omitted? (Draw the sequence.) Step (f) would then be called *charging by contact*. How does the sign of this charge on the sphere differ in the two cases? What would happen if in step (f) the negative charge on the rod were very large compared with the positive charge on the sphere?

**PROBLEM 21-4.** Repeat this series of pictures to illustrate the same sequence, starting, however, with a glass rod and a silk cloth.

**PROBLEM 21-5.** A metal rod held by means of an insulating handle may develop negative charges when rubbed with cloth. Redraw the sequence in Fig. 21-1 using a metal rod instead of a nonconductor.

Another interesting problem is presented by the action of an uncharged *nonconductor* in the presence of a nearby charged body. There is, of course, no flow of electrons on or in a true nonconductor, but we may imagine that the negative parts of each atom or molecule are pushed a little away from an external negative charge without being entirely severed from the positive parts; this corresponds to a distortion rather than a separation in the atoms or molecules of the material. Although such a small shift is negligible if it affects only a single atom, the total effect of billions of atoms is to *induce* charges, actually bound surface charges, on the two ends of the dielectric (see Fig. 21-2). As soon as the external charge is withdrawn, the induced surface charges usually also disappear; but until it is withdrawn the two

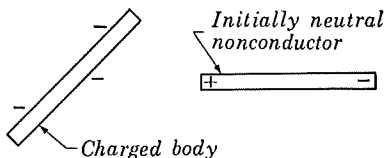


FIG. 21-2. Effect of induction on a nonconductor.

bodies in Fig. 21-2 will attract each other. We see therefore that a charged body attracts a previously neutral one, whether they are both conductors or not; and this explains the observation of the Greeks that a rubbed piece of amber attracts small pieces of lint.

**PROBLEM 21-6.** Redraw the sequence in Fig. 21-1 using a nonconducting sphere.

The exact degree of mobility of charges in different bodies depends a great deal on the experimental circumstances, and no really sharp division is possible. Cool, dry gas acts like a nonconductor, but very hot gases are good conductors. Perfectly pure water does not carry electric currents, but tap water containing certain impurities is a fairly good conductor and becomes an excellent one if a pinch of salt or any other substance (electrolyte) is added that separates in the solvent to form ions. The human skin is a fair conductor compared with clean glass, a fair nonconductor compared with metals.

The situation in conducting liquids (electrolytes) is more complex than in solids. The moving charges there are positive and negative ions produced when molecules of such substances as salts or acids become separated into oppositely charged parts on dissolving in the liquid. In heated or otherwise agitated gases, the currents consist of electrons as well as positive and negative ions.

**21-6 The electroscope.** In the discussion centering around Fig. 21-1 you may well have asked, How do we know what the charges are at every step? We certainly cannot tell the sign or even presence of an electric charge by just looking at the object.\*

One possible device for detecting the presence of a charge might be a tiny glass sphere at the end of a thin, long nonconductor. If we rub the sphere gently with silk we place on it a small positive charge. This sphere is then useful for exploring the neighborhood of another charged body: a strong repulsion indicates the presence of a like (positive) charge, and strong attraction the opposite.

However, we should soon find that we need a much better instrument—the leaf electroscope (Fig. 21-3). The metal rod *B* which is separated from the protecting case *D* by means of a good insulator *C*, is terminated at the top by a metal knob and at the bottom by very light leaves of gold or aluminum, *A*. If a charge (+ or -) is “placed” on the knob and allowed to distribute itself over the stem and leaves, then the charged leaves will fly apart by mutual repulsion. The same effect is seen if the neutral knob, instead of being charged by contact, is simply being approached by another charged body; then, by induction, there is a separation of charges in the knob and one type of charge is “driven” to the leaves, the other remains on the knob. Again the leaves deflect. In general, the deflection or divergence of the leaves indicates a presence of charge in the leaves, and we assume that the divergence of the leaves, which is a measure of the repulsion between them, also is a qualitative measure of the *quantity of charge*, to be defined more rigorously at once.

**PROBLEM 21-7.** Fig. 21-4 shows without text how an electroscope is charged by *induction*. Supply text for each picture; then repeat this series

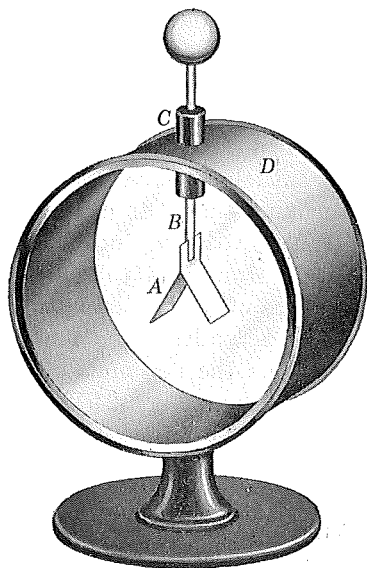


FIG. 21-3. The leaf electroscope.

\* Or by weighing, for the weight of the displaced electrons is inconceivably small compared with the bodies involved.

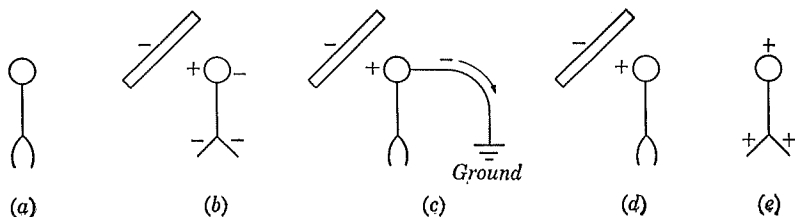


FIG. 21-4. Charging an electroscope by induction.

of pictures for the case where an initially positively charged rod is used. (Note: Only the sign, not the quantity, of the *excess* charges is indicated.)

A good electroscope, once charged, will hold its charge extremely well, as seen by the steady deflection of the leaves; of course, the insulation at *C* must be carefully chosen, and all traces of conducting films of moisture or grease must be absent. If, however, the air surrounding the knob is heated or is made conducting in some other way, the charge leaks off and the leaves collapse. From the rate of collapse one may judge the strength of a source of x-rays or of radioactive emission, for both types of high-energy radiation "ionize" the gases through which they pass. Hence electroscopes are important instruments in modern research, although the details of construction may differ widely from Fig. 21-3. For example, one type of rugged electroscope, externally only as small and slim as a fountain pen, nowadays is carried in the pockets of many research workers who must expose themselves to nuclear radiations; the electroscope is charged at the beginning of the day, and afterwards a periodic check of its deflection reveals whether the carrier has had more than the experimentally known safe dosage of high-energy radiation.

Electroscopes of some form have been prominent in some of the most important discoveries of modern physics. One of the first clues of *cosmic rays* came when it was found that, in time, even the very best and the most carefully guarded electroscopes steadily lose some of their charge; the residual ionization of the air around the knob or leaves persisted even if the electroscope was shielded by thick layers of brick or metal.

**PROBLEM 21-8.** Explain the following simple phenomena or devices in terms of the discussion of electrostatics: Walking across a carpet is followed by a shock when the hand touches a metal object. Violent motion of moist air or raindrops is followed by lightning. Gasoline trucks often drag chains, and cars are brushed by metal strips as they cross a toll bridge.

**21-7 The quantity of charge.** Figure 21-5 is a schematic representation of one of the famous experiments by the great English experimenter Michael Faraday (1791-1867), in which he (a) connected a metal can (actually an ice pail) with the knob of a discharged electroscope; then a charged metal object (say a positively charged sphere) on an insulating handle was lowered into the pail, (b), yielding a constant deflection of the leaves no matter where in the pail the charged sphere was held. In (c) the sphere touched the inside of the pail—no effect on the leaves is observed; in (d) the sphere was withdrawn, found to be uncharged, and the leaves, which still remain distended, were found to be positively charged. Working back from these last results, Faraday concluded that the amounts of  $+$  and  $-$  charges separated by induction are exactly equal, and that a hollow conductor such as this pail normally carries all *excess* charges on the outside surface.

**PROBLEM 21-9.** (a) Redraw Fig. 21-5, showing the sign of any excess charges in place of the question marks. Explain your choices. (b) Show how Faraday's conclusion follows from these results. (c) What if the sphere initially had been negatively charged?

Apart from the proof of equality of  $+$  and  $-$  induced charges, Faraday's experiment provided a very useful tool. If we lower a body  $A$  into the (discharged) can, the electroscope leaves diverge; if we withdraw  $A$  again without touching the can, the leaves collapse completely. We may decide to make the deflection of the leaves *a measure of the quantity of charge*. For example, some  $+$  charged

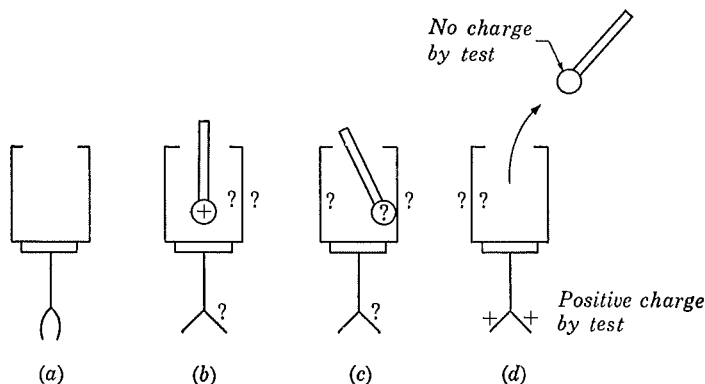


FIG. 21-5. Faraday's "ice-pail" experiment.

body  $A$  produces perhaps a  $5^\circ$  deflection of the leaves; we arbitrarily and provisionally call the charge on that body a *unit + charge*. Another body  $B$  is now charged positively until, by itself, it also deflects the leaves  $5^\circ$ . Both  $A$  and  $B$  are now held in the can together, and we notice a deflection of, say,  $7^\circ$ ; this deflection corresponds to a charge of 2 units, by definition. We go on to calibrate the electro-scope step by step so that in the future we can always determine the relative quantity of charge on any two bodies by the deflections of the leaves produced on lowering the two bodies in turn into the discharged can. Parenthetically, we might note with satisfaction that a unit + charge (a  $5^\circ$  deflection) and a unit - charge (also  $5^\circ$ ) when lowered together, do not deflect the leaves at all, i.e., that equal and opposite charges "cancel" when brought together.

This process of calibration, analogous in some respects to the comparison of masses by means of a calibrated spring balance, has some serious drawbacks (which?), and it is not used in actual practice to define quantity of charge. Another method suggests itself, necessarily also based on the observable forces that charges exert on one another.

**21-8 Coulomb's law of electrostatics.** Place some electric charges on two very small metal spheres; the relative amounts of charge, by measurement with the can arrangement, are  $q_1$  and  $q_2$  in magnitude (e.g.,  $q_1 = -2$  units,  $q_2 = +2$  units). Both spheres—so small and so far apart that we may call them *point* charges—are suspended in vacuum, and are so far away from other materials that the observed forces on the charged spheres are in effect only the electric force  $F$  of  $q_1$  on  $q_2$  and the reaction of  $q_2$  on  $q_1$ . *By experiment*,  $F \propto 1/r^2$ , where  $r$  is the distance between the two point charges. Also,  $F \propto q_1$  ( $q_1$  may be increased to  $-3, -4, -5, \dots$  units by recharging and referring to the electro-scope calibration); and  $F \propto q_2$ . In short,  $F \propto q_1 q_2 / r^2$ , quite analogous to the law of gravitation. But since we have not yet committed ourselves permanently to any specific units of charge, we may write

$$F = \frac{q_1 q_2}{r^2} \quad (21-1)$$

and thereby we redefine the unit of electric charge as follows: one electrostatic unit of charge (1 esu of charge) is the quantity of a point charge which, 1 cm away from an equally large point charge, exerts on it 1 dyne of force. Note that  $F$  is a *vector* lying along a line through  $q_1$  and  $q_2$ .



PROBLEM 21-10. (a) How do we know when  $q_1$  and  $q_2$  are each 1 esu? (b) Show that 1 esu of charge =  $1\sqrt{\text{dyne-cm}^2}$ . (c) From accurate experiments it is known that each electron carries about  $4.8 \times 10^{-10}$  esu of negative charge. The neutral hydrogen atom has only one such electron, about  $0.5 \times 10^{-8}$  cm from its positively charged central nucleus. What is the electric force holding the electron to the nucleus? If the mass of an electron is approximately  $10^{-27}$  gm and of a hydrogen nucleus is about  $2 \times 10^{-24}$  gm, what is the magnitude of the gravitational force compared with the electrostatic force? (d) How many electrons make up  $-1$  esu of charge?  $+1$  esu? (e) How many electrons must be placed on each of two spheres 10 cm apart to cause each sphere to exert a force of 1 dyne on the other?

The conditions leading to our definition of quantity of charge seem practically absurd, but it is not impossible to approximate the requirements to within known and calculable limits, so that Eq. (21-1) can, in fact, be regarded as practically useful and effectively confirmed. The actual details of accurate measurements, which are very similar to those involved in the confirmation of the gravitational force equation, were first published by Coulomb in France in 1785; therefore Eq. (21-1) is called *Coulomb's law*,\* and 1 esu of charge is often called 1 *statcoulomb*.

It has been found convenient to define a larger unit of electric charge, the *coulomb* (1 coulomb =  $3 \times 10^9$  statcoulombs very nearly), but since forces are to be expressed in dynes and distances in centimeters, our Eq. (21-1) applies in that form *only* if the charges  $q_1$  and  $q_2$  are converted to statcoulombs.

By experiment, it has also been found that if air instead of a vacuum is placed between  $q_1$  and  $q_2$ , the force these charges experience is sensibly unchanged. However, if the space between and around the two charges is completely filled with another homogeneous isotropic dielectric such as paraffin or pure water, the force experienced by each charge is reduced by a factor of about  $\frac{1}{2}$  and  $\frac{1}{81}$  respectively. This is not because Coulomb's law is there invalid, but because in the presence of  $q_1$  and  $q_2$  the charges in the dielectric separate and in turn exert forces of their own upon  $q_1$  and  $q_2$  that tend to counteract the direct effects of  $q_1$  on  $q_2$  and  $q_2$  on  $q_1$ . However, this is a complex situation, and we shall usually have little occasion to talk about any cases except that of charges *in vacuo*, for which Eq. (21-1) describes accurately the total force actually observed.

---

\* The accuracy of the statement  $F \propto 1/r^2$  in Coulomb's law has recently been confirmed indirectly to within 1 part in one billion.

## 21-9 Applications of Coulomb's law.

*Example 1.* Three point charges in vacuum,  $q_1$ ,  $q_2$ , and  $q_3$ , are situated at the corners of a right triangle at distances shown in Fig. 21-6(a), and  $q_1 = +10$  esu,  $q_2 = +\frac{1}{2} \times 10^{-8}$  coulomb,  $q_3 = -2 \times 10^{-9}$  coulomb. Find the net force  $F$  on  $q_1$ .

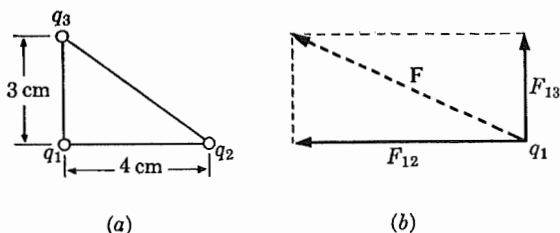


FIG. 21-6.

*Solution.* The net force on the point charge is the *vector sum* of the forces due to each of the other two point charges. If  $F_{12}$  is the force on  $q_1$  due to  $q_2$ ,  $F_{12} = q_1q_2/r_{12}^2$ ; but all charges are to be expressed in esu, therefore we convert:

$$q_2 = +\frac{1}{2} \times 10^{-8} \text{ coul} = +\frac{1}{2} \times 10^{-8} \text{ coul} \times \frac{3 \times 10^9 \text{ statcoul}}{\text{coul}} = +15 \text{ statcoul},$$

$$q_3 = -2 \times 10^{-9} \text{ coul} = -2 \times 10^{-9} \text{ coul} \times \frac{3 \times 10^9 \text{ statcoul}}{\text{coul}} = -6 \text{ statcoul}.$$

$$\text{Now } F_{12} = \frac{(+10 \text{ statcoul})(+15 \text{ statcoul})}{(4 \text{ cm})^2} = +\frac{150}{16} \text{ dynes}.$$

Since  $q_1$  and  $q_2$  are of the same sign,  $F_{12}$  is a force of repulsion, acting on  $q_1$  to the *left*. (We can always identify a force of repulsion by the + sign in the result, a force of attraction by the - sign.)

The other of the two forces acting on  $q_1$  is  $F_{13}$  due to  $q_3$ .

$$F_{13} = \frac{q_1q_3}{r_{13}^2} = \frac{(+10 \text{ statcoul})(-6 \text{ statcoul})}{(3 \text{ cm})^2} = -\frac{60}{9} \text{ dynes},$$

a force of attraction on  $q_1$ , acting *upward* in the sketch.

The *net* force  $\mathbf{F}$  on  $q_1$  is the vector sum of  $F_{12}$  and  $F_{13}$  [Fig. 21-6(b)] obtained in the usual way by vector addition. (Carry out.)

PROBLEM 21-11. Use the data in the above example to find the net force (magnitude and direction) on  $q_2$ .

PROBLEM 21-12. Plot a graph of relative force on a point charge  $q'$  vs. distance from a second charge  $q$  if there are no other charges present.

*Example 2.* To find the electrostatic net force on a point charge  $q$  in the neighborhood of a large charged body, we must add vectorially the several forces exerted individually on  $q$  by each of the point charges distributed on the large body. Evidently this is tedious, if not impossible, in a general case with the mathematical methods at our command, but there are two special cases which come up so often in our later studies that we might now at least look at the obtainable results.

(a) The net force  $F$  on a single small point charge  $q$  in air or vacuum *outside* a uniformly charged *spherical* body is given by  $F = q \cdot Q/r^2$ , where  $Q$  is the net charge on that spherical body, and  $r$  is the distance between  $q$  and the center of the sphere. Note the similarity between this conclusion and Newton's results for the gravitational attraction exerted by a spherical mass on a small point mass. There, too, we heard that the effect of a (homogeneous) spherical body is the same as if the bulk of the sphere were shrunk to a point at its center.

(b) The net force on a single small point charge is constant in direction and magnitude anywhere inside the air gap between two oppositely charged large parallel *plates*, an arrangement called a condenser or capacitor. (Note that the force on a charge is *not* constant along the line between two opposite *point* charges; cf. problem 21-12.)

PROBLEM 21-13. Compare Newton's law of universal gravitation as applied to two small point masses, and Coulomb's law of electrostatics applied to two point charges; consider the similarities and outstanding differences. What conclusions do you draw from the repeated appearance of the inverse-square law in the various branches of physics? (Do not assume that a dogmatic answer to this question exists.)

**21-10 The electrostatic field.** If we attach a small body to the end of a thin nonconducting stick and give it a positive charge, say +1 statcoulomb, the body (a test point charge) will experience a force  $F$  when placed anywhere in the vicinity of other electric charges. We may use this probe to explore every point in space, and to find the electric force (in magnitude and direction) at every point. We could populate this whole space mentally with arrows to a given

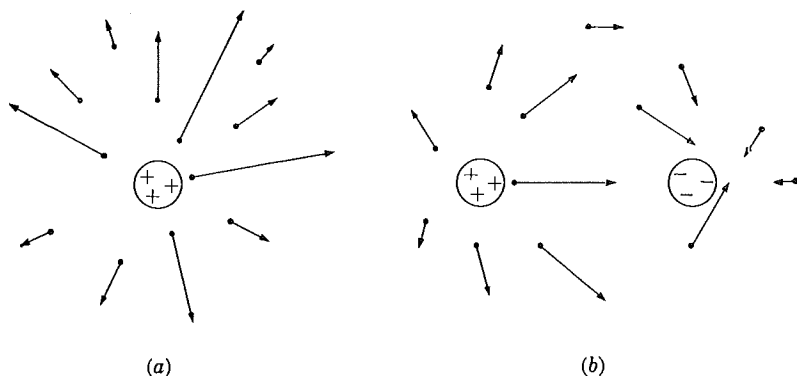


FIG. 21-7. Force vectors on a small positive test charge ( $q$  statcoulomb) at selected points near fixed charges.

scale to show the situation at every point at a glance. In Fig. 21-7 a few such electric force vectors have been represented in the neighborhood of simple charge distributions, namely (a) around a positively charged sphere, (b) around two oppositely charged spheres. Obviously the force vectors will become negligibly small far away from these charges, but even in their immediate neighborhood we realize that there must be an infinity of such vectors, one for each point of space. It is inconvenient that these force vectors depend on two factors, i.e., on both the charge  $Q$  that sets up the action and the charge  $q$  that is acted upon. Therefore another very useful concept has been defined which, by a standardization of the test charge used to explore the action of  $Q$ , is independent of  $q$ . This is called the *electric intensity at  $P$*  or *electric field strength at  $P$* , and is directly defined as the electrostatic force per positive unit charge placed at that point  $P$ . The letter  $E$  is usually assigned to this concept. By definition,

$$\mathbf{E} \text{ at } P = \frac{\text{net force } \mathbf{F} \text{ on positive charge } q \text{ at } P}{q}, \quad (21-2)$$

usually expressed in units of dynes/statcoul. The arrows in Fig. 21-7, if they referred to forces *per unit charge*, now represent directly the vector  $\mathbf{E}$  at each selected point. *The totality of such  $E$ -vectors in a given region we call "the electrostatic field" in that region.*

This is a loose definition of one of the most fruitful concepts of physics (gravitational field, magnetic field, etc., are similarly conceived).

Note three important characteristics of the concept "electric intensity." (1) The *magnitude* of  $E$  at a point  $P$  may be determined by measuring the force  $F$  on a test charge of, say,  $+15$  statcoul placed at  $P$ , and dividing this force  $F$  by  $15$  statcoul, since  $E$  is the *ratio* of force by charge. If  $30$  statcoul had been chosen for the test charge,  $F$  would have been twice as large, but the ratio  $E$  at that point comes to the same value no matter how large the test charge.\* The magnitude of  $E$  is determined by the charges that *set up* the field, and not by the small test charges used to explore its strength. In fact, that is the main usefulness of the concept.

(2) The *direction* of  $\mathbf{E}$  at  $P$ , by convention, is the same as the direction of the force on a *positive* test charge at  $P$ . We might equally well have chosen the opposite convention, but once we agree on the present one we must stick to it consistently.

(3) The *units* of  $E$  are the units of force/charge, usually dyne/statcoulomb (dyne/coulomb is also a possible choice); but it will save a great deal of bother if we express our charges for the time being in statcoulombs.

*Example 3.* (a) The force  $F_P$  on a  $+4$  statcoulomb test charge at a point  $P$ ,  $10$  cm from a small charged sphere, is  $12$  dynes toward the sphere, by test. Find the electric intensity  $E_P$  at  $P$ .

*Solution.*  $E_P = 12 \text{ dynes}/+4 \text{ statcoul} = 3 \text{ dynes/statcoul}$  (toward sphere).

(b) Now find the total charge  $Q$  on the sphere.

$$\textit{Solution.} \quad F_P = 12 \text{ dynes} = \frac{Q \cdot (+4 \text{ statcoul})}{100 \text{ cm}^2};$$

$$\text{therefore} \quad Q = \left( \frac{12}{4} \times 100 \right) \text{ statcoul},$$

and  $Q$  is *negative* since it attracted a  $+$  test charge.

The defining equation  $E = \frac{\text{force on positive test charge}}{\text{magnitude of test charge}}$  becomes particularly simple if the field-producing charge  $Q$  is a point charge at a distance  $r$  from the location of test charge  $q$ , or if  $Q$  is uniformly distributed on any sphere whose center is  $r$  cm from  $q$ . In either case (see Fig. 21-8),

$$E_P = \frac{F_P}{q} = \frac{\left( \frac{Q \cdot q}{r^2} \right)}{q} = \frac{Q}{r^2} \quad (21-3)$$

\* But the test charge still must not be so large that it induces perceptible charge rearrangements in its neighborhood.

But for all other cases, the value of  $E$  at point  $P$  must be found by experiment from the defining equation (force/charge), or else by adding vectorially the contributions to  $E$  at  $P$  of each point charge  $Q$  at its particular distance  $r$ , each contribution calculated by Eq. (21-3). We have here neither the tools nor the need for carrying out this mathematically advanced exercise.

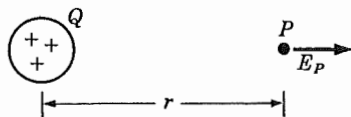


FIG. 21-8.

**PROBLEM 21-14.** At a distance of 100 cm from the center of a uniformly charged large sphere the electric intensity is 4 dynes/statcoul. What is the force on a point charge of  $10^{-9}$  coulomb at that distance from the sphere, and what is the total charge in coulombs on the sphere?

**21-11 Lines of force.** To visualize the electric field around a charged body we must try mentally to supply each and every point in space with an arrow indicating the force per unit charge at that point; this is obviously a theoretically feasible but practically impossible task. As an aid to our limited powers of imagination we may picture the field in another way.

We know that if we bring a small positive test charge  $q$  into an electric field created by some other charges,  $q$  will experience repulsive forces from  $+$  charges, attractive forces to  $-$  charges. In Fig. 21-9 are shown the paths of a small positive test charge  $q$  released at various points in the neighborhood of some other charges. All forces except electric ones are neglected, and the inertia of the body carrying  $q$  is

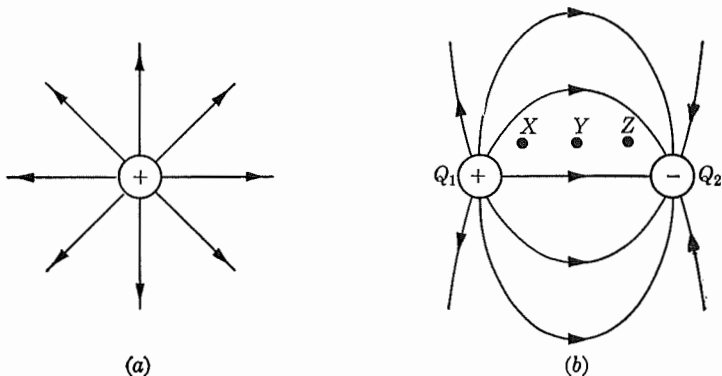


FIG. 21-9. Lines of force, (a) in the neighborhood of a single positive charge, (b) around equally large point charges of opposite sign.

also negligible. Therefore the direction of travel (indicated by the arrow on each line) follows the direction of the *force* acting on  $q$  at every point along the path, and we may call these lines the *lines of force*. At every point along each line, the force acting on  $q$  is tangential to the line, but the line only informs us of the direction of the force vector on the positive charge, not of its magnitude.

**PROBLEM 21-15.** Invent a modification of this idea to present at one glance both direction and magnitude at every point along the path.

Clearly it is possible in every case to draw infinitely many different lines of force around a charged body. Every point in space lies on a line of its own; for example, you can readily imagine (and draw now) a curve from  $Q_1$  to  $Q_2$  through point  $X$  in Fig. 21-9(b). But even without filling in all possible lines, we get a good feeling for the state of affairs by looking at only a few such lines. As Fig. 21-9(a) informs us, a crowding of the lines of force means numerically a relatively stronger field (near the charged body), a larger spread means numerically a relatively smaller field strength  $E$ . Thus point  $X$  is in a stronger field in Fig. 21-9(b) than point  $Y$ ; and in this direct relation, so pictorial and helpful to our imagination, lies much of the usefulness of this concept of lines of force.

**PROBLEM 21-16.** (a) Note that a line of force, starting at a positive charge, either goes to infinity or terminates on a nearby negative charge. Why? (b) Can lines of force intersect each other? (c) Draw some lines of force (i) around two positive charges placed like  $Q_1$  and  $Q_2$  in Fig. 21-9(b), (ii) between a parallel plate condenser charged oppositely on the two plates. (d) Is the electric field around a charged body *real*? Are the lines of force real?

**PROBLEM 21-17.** It is a fact that if a conductor carries electric charges which have found their "equilibrium" position (steady state), the lines of force join the conductor perpendicularly at every point on the surface. Explain why. (*Hint:* Consider what would happen to the charges on the surface if the lines of force were oblique instead of normal.)

**21-12 Electric potential difference—qualitative.** We turn now to another concept which is of fundamental importance in electricity—electric potential difference, for which the symbol  $V$  will be used.

The word *potential* reminds us at once of potential energy, defined in mechanics. In fact, the new concept also refers to work done by virtue of a change of position (of charged bodies), although it is now work done by or against electric forces rather than gravitational or elastic ones. But the analogy is quite deceptive in many respects

and is better not invoked at all. For a definition we rather choose the following.

The potential difference  $V_{AB}$  between any two points  $A$  and  $B$  is the ratio obtained when the work done in moving a positive electric charge  $q$  from  $B$  to  $A$  is divided by the magnitude of the charge  $q$ ;  $V_{AB}$  = work done per unit charge moved, in ergs/statcoul or in other appropriate units such as joules/coul. (By definition, 1 erg/statcoul = 1 statvolt; or, in the so-called "practical" system of units, 1 joule/coul = 1 volt.)

For example, refer again to Fig. 21-9(b). When we attempt to carry a positive charge  $q$  from  $Y$  to  $X$ , our muscles must do work against the repulsion from  $Q_1$  on the left and the attraction to  $Q_2$  on the right. The net work done per unit charge can be found experimentally (suggest a way), and turns out to be independent of the exact path chosen between  $Y$  and  $X$ , just as the net work done in raising a weight depends on the total displacement rather than on the exact path taken.

By convention, we say that  $X$  is at a more positive potential than  $Y$  or  $Z$  because our muscles or some other nonelectric forces must do work on a positive charge to carry it from  $Y$  or  $Z$  to  $X$ . On the other hand,  $Z$  is at a more negative potential with respect to  $Y$  or  $X$  because on carrying a positive charge to  $Z$  from  $Y$  or  $X$  all the work is done by the electric field instead of our muscles. In brief, the relative positive and negative potential at a point is recognized by the behavior of a positive charge which, "on its own," tends to go *away* from a point at relatively *positive* potential and toward one at relatively *negative* potential.

While in this manner the words "potential difference between  $A$  and  $B$ " are easily given operational significance, the frequently encountered phrase "the potential at point  $A$ " is not so clear, because no mention is made in it of the starting or reference point. *By convention*, the *implied* reference point is some infinitely distant uncharged body or, in practice, for example, the surface of the earth ("electrical ground") at some distance below point  $A$ . Therefore if we hear "the potential is positive at point  $A$ ," we know that a positive test charge brought to point  $A$  will experience a net force of repulsion and will try to escape to ground if a conductor is stretched from point  $A$  to ground (see Fig. 21-10). Similarly, a point  $A'$  is at a negative potential if a positive charge is readily attracted from ground toward that point  $A'$ . The numerical value of the potential at a point  $A$  (no second reference point given) is then simply the potential difference between  $A$  and any point on the ground, or the



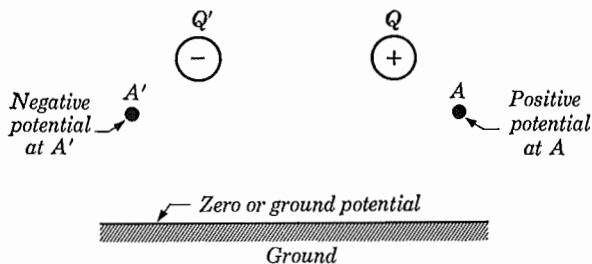


FIG. 21-10.

work done per unit charge on taking a positive charge from ground to point  $A$ . For all our purposes, then, we may always replace the words "potential at a point" by "potential difference between that point and ground."

**PROBLEM 21-18.** What is the operational meaning of the phrase "a point  $P$  has a negative potential of 1 erg/statcoul"? How many joules/coul is that? (*Answer:* 300 joules/coul. Remember this rate of exchange—it is frequently invoked.)

**PROBLEM 21-19.** Decide whether the potential is  $+$ ,  $-$ , or zero at points  $X$ ,  $Y$ , and  $Z$  in Fig. 21-9(b), providing  $Q_1 = Q_2$  numerically but is of opposite sign.

**PROBLEM 21-20.** At electrostatic equilibrium (no current) the potential difference between any two points on the surface of a conductor is zero. Explain why.

**PROBLEM 21-21.** In Fig. 21-10 there will also be points between  $Q$  and  $Q'$  where the potential is zero. Indicate a few. (*Hint:* The potential at any point  $P$  with respect to ground is zero if no work is done moving a charge from ground to  $P$ . Furthermore, no work is done moving always perpendicular to lines of force; why?)

**21-13 Potential difference—quantitative.** So far we have developed only the qualitative and the experimental aspects of the concept of potential difference; now we ask the inevitable question, How can one *calculate* and predict the potential difference between two points?

Consider the simplest possible case, the potential difference  $V_{AB}$  between points  $A$  and  $B$ , in the vicinity of a charged sphere  $Q$  and away from all other charges (Fig. 21-11). Let us assume the charge  $Q$  is positive. Point  $A$  is therefore at a higher positive potential than  $B$ , since one has to supply work to transport a small positive charge  $q$  from  $B$  to  $A$  against the repulsion of  $Q$ .  $V_{AB}$  is this work per unit charge; and if the force on  $q$  between  $B$  and  $A$  were constant

(say  $F$ ) the work in this simple case would be  $F \cdot (r_B - r_A)$ ,  $V_{AB}$  being consequently  $F(r_B - r_A)/q$ . But the force on  $q$  is not at all constant during its transit from  $B$  to  $A$ . In vacuum or air the force changes from  $Qq/r_B^2$  at  $B$  to  $Qq/r_A^2$  at  $A$ , and not linearly either. Frankly, we do not know how to calculate the exact work done under such circumstances without calculus, but we can make an instructive approximate calculation of  $V_{AB}$  which will point to the simple result

$$V_{AB} = \frac{Q}{r_A} - \frac{Q}{r_B}. \quad (21-4)$$

*Derivation.* To find an approximate value for the work done on  $q$  from  $B$  to  $A$  we mentally subdivide the path  $BA$  into very small parts, each of length  $\Delta s$ . The force does not change much during a small displacement  $\Delta s$ . For example, on going from  $B$  at distance  $r_B$  to a near point at distance  $r_1$  from  $Q$  (whence  $\Delta s = r_B - r_1$ ), the average force is nearly  $Qq/r_B^2$  or also  $Qq/r_1^2$  or  $Qq/r_B r_1$ ; the smaller  $\Delta s$ , the more nearly true this is. But the work  $\Delta \mathfrak{W}_1$  done during this small displacement,  $\Delta s$ , is simply the average force  $\times$  displacement (forces and displacements being here parallel).

$$\Delta \mathfrak{W}_1 = \frac{Qq}{r_B r_1} \Delta s = \frac{Qq}{r_B r_1} (r_B - r_1) = Qq \left( \frac{1}{r_1} - \frac{1}{r_B} \right).$$

On proceeding toward  $A$  over the next small interval to a near point at distance  $r_2$  from  $Q$ , i.e., through a distance  $(r_1 - r_2)$ , the work done is

$$\Delta \mathfrak{W}_2 = Qq \left( \frac{1}{r_2} - \frac{1}{r_1} \right),$$

similarly obtained. The total work on  $q$  from  $B$  to  $A$ , the sum of such contributions for  $n$  small intervals, is

$$\begin{aligned} \mathfrak{W}_{\text{total}} &= \Delta \mathfrak{W}_1 + \Delta \mathfrak{W}_2 + \cdots + \Delta \mathfrak{W}_n \\ &= Qq \left( \frac{1}{r_1} - \frac{1}{r_B} \right) + Qq \left( \frac{1}{r_2} - \frac{1}{r_1} \right) + \cdots + Qq \left( \frac{1}{r_A} - \frac{1}{r_n} \right) \end{aligned}$$

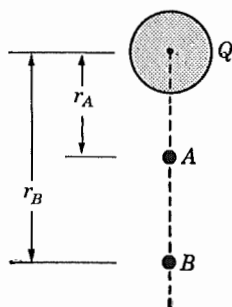


FIG. 21-11. The potential difference between points  $A$  and  $B$  near a uniformly charged sphere  $Q$  is  $V_{AB} = Q/r_A - Q/r_B$ .

$$\begin{aligned}
 &= Qq \left( \frac{1}{r_1} - \frac{1}{r_B} + \frac{1}{r_2} - \frac{1}{r_1} + \cdots + \frac{1}{r_A} - \frac{1}{r_n} \right) \\
 &= Qq \left( \frac{1}{r_A} - \frac{1}{r_B} \right), \text{ all other terms canceling.}
 \end{aligned}$$

By definition,  $V_{AB} \equiv \frac{W_{\text{total}}}{q} = \frac{Q}{r_A} - \frac{Q}{r_B}$ ; Q.E.D.

This last result applies only to the region (in air or vacuum) around a single point charge  $Q$ , or to a single sphere on which  $Q$  charges are uniformly distributed. An important consequence of Eq. (21-4) becomes apparent if we let  $r_B$  become very large. As  $r_B \rightarrow \infty$ ,  $V_{AB}$  now means the potential difference of point  $A$  with respect to an infinitely distant uncharged body (e.g., to ground);  $V_{AB}$  is now "the potential at a point  $A$  which is  $r_A$  cm from a single point charge  $Q$  or a uniformly charged sphere." In this case, since  $1/r_B = 1/\infty = 0$ , let us write simply  $V_A$  instead of  $V_{AB}$ , and we have

$$V_A = \frac{Q}{r_A}. \quad (21-5)$$

**PROBLEM 21-22.** In problem 21-10 we considered the force at  $0.5 \times 10^{-8}$  cm from a nucleus of a hydrogen atom. Now find the potential at that distance. Also find the potential difference between two points  $0.5 \times 10^{-8}$  cm and  $2 \times 10^{-8}$  cm from the nucleus respectively.

**PROBLEM 21-23.** Find the potential at a point halfway between  $Q_1$  and  $Q_2$  in Fig. 21-9(b) if  $Q_1 = +3$  statcoul,  $Q_2 = -5$  statcoul, and the distance between them is 10 cm.

**21-14 Uses of the concept of potential.** There are several interesting features of Eq. (21-5). (a) If the sign of the charge  $Q$  is used along with its numerical value,  $V_A$  automatically appears with its correct  $+$  or  $-$  sign. (b) The units of  $V_A$  appear to be statcoul/cm, whereas previously we decided they should be ergs/statcoul. This is not really a paradox, however, since  $1 \text{ erg/statcoul} = 1 \text{ statcoul/cm}$  (prove by recalling that  $1 \text{ statcoul} = \sqrt{1 \text{ dyne-cm}^2}$ ). (c) Like the concept of work itself, potential is a scalar; therefore simple algebraic addition gives the total potential at any point in the vicinity of more than a single point charge. Hence a point equally far from equally strong opposite charges, like point  $Y$  in Fig. 21-9(b), is at zero potential with respect to ground (but the *force* is *not* zero there! Is this a contradiction?) (d) In our derivation there was no restriction on how close to the charged sphere the point  $A$  may be for the equation  $V_A = Q/r_A$  to hold. Actually, point  $A$  may even be right

on the surface of the sphere, in which case the potential of the sphere,  $V_s$ , is the ratio of its charge  $Q$  to its radius  $r_s$ , i.e.,  $V_s = Q/r_s$ . Thus 1 coulomb of charge deposited on a 2-cm sphere ( $r_s = 1$  cm) raises the sphere to  $3 \times 10^9$  statcoul/cm or  $3 \times 10^9$  ergs/statcoul potential; recalling that 1 erg/statcoul = 300 joules/coul  $\equiv$  300 volts, we see that the sphere would be at  $9 \times 10^{11}$  volts potential. It is clear that relatively small bodies can easily achieve large potentials if supplied with enough charge; a rubber or glass rod can quickly be given enough charge by friction to develop several thousand volts on a dry day.

The usefulness of the concept *electric potential* may be further illustrated by two other instances:

(e) We can now quickly calculate the potential energy of a point charge  $q$  at distance  $r_A$  from a point charge  $Q$ . The potential at that distance is  $V_A$ ; by definition this is the potential energy of a charged body *per unit charge*. Therefore the total potential energy of  $q$  at  $A$  is the product  $V_A \times q$ . If, for example,  $Q$  refers to the nucleus of a hydrogen atom and  $q$  refers to its electron at  $r_A = 0.5 \times 10^{-8}$  cm, the potential energy of the electron is

$$\begin{aligned} \text{PE} &= \frac{Q}{r_A} \times q = \frac{4.8 \times 10^{-10} \text{ statcoul}}{0.5 \times 10^{-8} \text{ cm}} \times (-4.8 \times 10^{-10} \text{ statcoul}) \\ &= -4.6 \times 10^{-11} \text{ erg.} \end{aligned}$$

(What is the meaning of the negative sign?)

(f) We have found before that the electric force  $F_E$  on a charge between two equally and oppositely charged plates is constant. As a consequence, the electric intensity  $E$  in such a gap is also constant. The work that must be done *per unit charge* to transport a small positive test charge  $q$  from the negative plate a distance  $s$  straight across the gap to the positive plate (Fig. 21-12) is, by definition,  $V$  (its value is read off most conveniently from the label of the battery that might be used to provide the potential difference). In this situation, the work that must be done on  $q$  against the electric forces is work per unit charge  $\times$  total charge, or

$$F_E \times s = V \times q.$$

Or transposing, and noting that  $F_E/q = E$ ,

$$E = V/s. \quad (21-6)$$

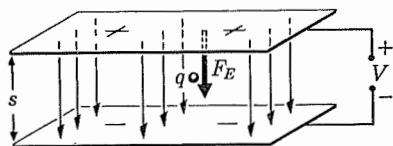


FIG. 21-12. The electric force  $F_E$  on a small charged body  $q$  between two oppositely charged plates.

If  $V$  is in ergs/statcoul (= statvolts) and  $s$  is in cm,  $E$  is in dynes/statcoul; if  $V$  is in volts and  $s$  is in cm,  $E$  is in volt/cm. You may prove quickly that  $1 \text{ volt/cm} = 10^7 \text{ dynes/coul} = 1/300 \text{ dyne/statcoul}$ .

The importance of Eq. (21-6) is that since  $s$  is easily measured and  $V$  is known from the battery used,  $E$  can be readily calculated at all points of the space between the plates.

**PROBLEM 21-24.** Two parallel horizontal plates are 2 cm apart in vacuum, and are connected to a supply providing a 10,000-volt potential difference. An electrically charged oil drop, which from other measurements appears to have a mass of  $10^{-4}$  gm, is seen to be suspended motionlessly between the plates. What is the electric force on that oil drop? What is its charge  $q$ ? What charge must exist on the drop to account for its vertical motion in the electric field with an acceleration of  $2 \text{ cm/sec}^2$  up?  $10 \text{ cm/sec}^2$  down?

**PROBLEM 21-25.** A Van de Graaff generator is essentially a metallic sphere on an insulating stand, charged to a high electric potential, which is used to repel and accelerate charged atomic particles along an evacuated tube to reach extremely high speeds; when they crash into a "target" they will have an energy suitable for "atom smashing." If we wish to use in this accelerator a sphere 10 ft in diameter at 2,000,000 volts, how much charge must be supplied? (Watch the units.) What might be the practical limits to reaching much higher potentials with such a device?

**PROBLEM 21-26.** If an electron is repelled from a point near the surface of this same Van de Graaff generator and is allowed to accelerate freely, what will be its final kinetic energy as it hits a target far from the generator? What will be its final speed? (Mass of electron =  $9.11 \times 10^{-28}$  gm, charge =  $4.8 \times 10^{-10}$  statcoul.) What was the potential energy of the electron just as it began to leave the surface of the sphere?

**PROBLEM 21-27.** How much kinetic energy (in joules and in ergs) does an electron gain when it freely passes through a potential difference of 1 volt? [This quantity of energy has been given the unfortunate name *1 electron volt* (1 ev), an important unit of energy in atomic physics.]

**21-15 Electrochemistry.** The history of electricity shows that the fundamental notions of charge, conductor, insulator, induction, and the like were acquired only slowly and sometimes after lengthy struggles. It was not until about 1800, the year of Alessandro Volta's discovery of the voltaic cell, that the subject of electricity really opened itself to widespread and fruitful investigations touching on the structure of matter. One of the first consequences was the discovery of the spectacular decomposing effect of currents on certain liquids. Electric current proved to be a tool capable of

breaking up into their elements dissolved or molten materials whose molecules previously had resisted separation by chemical or physical means. It was by this process, called electrolysis (Section 21-5), that Davy in 1807 broke down potash and soda, until then considered to be elements, and obtained from them the metals potassium and sodium.

While the study of electric currents led physicists to the subject of electromagnetism, which is part of our next consideration, further research on electrolysis itself led chemists to the atomic hypothesis, by another door than Dalton's. In a systematization and extension of Davy's methods, Michael Faraday proved that the passage of current through a liquid (e.g., a diluted acid or a salt solution) would dissociate the liquid at the two electrodes in a regular and accurately repeatable manner. The mass of an element liberated from the liquid, either in the form of gas bubbles (as for hydrogen or chlorine) or as metal deposited at or near the electrode (e.g., copper), is proportional both to the molecular weight of the material freed and to the quantity of charge transferred, but inversely proportional to the valence of the material liberated. For example, charge passing through a molten sample of common salt (NaCl) caused the sodium metal to collect at the negative electrode (the *cathode*, Fig. 21-13), and liberated chlorine gas bubbles at the positive electrode (*anode*). The mass of metallic

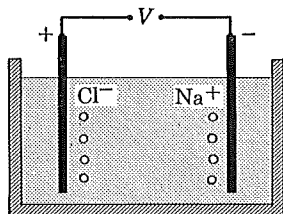


FIG. 21-13. Migration of ions to electrodes during electrolysis.

sodium is found to be about  $8 \times 10^{-14}$  gm per statecoulomb of charge so transferred, and the mass of chlorine gas about  $12 \times 10^{-14}$  gm. It follows that in order to obtain 1 gram-atomic-weight of Na (namely, 23 gm), one needs  $23 \text{ gm} / (8 \times 10^{-14} \text{ gm/statecoul})$  or 96,500 coul; at the same time, the transfer of this quantity of charge (occasioned, say, by a 26.8 ampere current lasting one full hour) will also release 1 gram-atomic-weight of chlorine (35.5 gm) at the other electrode. The transfer of twice that charge yields twice the respective amounts of these monovalent substances. If we now electrolyze water, 96,500 coul will release 1 gm of hydrogen but only 8 gm of oxygen, for oxygen has the valence 2, and the experimental law of electrolysis teaches that a quantity of charge capable of releasing 1 gram-atomic-weight of a *monovalent* element will produce only  $\frac{1}{2}$  gram-atomic-weight of a *divalent* element. In equation form these experimental

facts (essentially Faraday's law of electrolysis in modern terms) can be summarized as follows:

Mass of material liberated at one electrode (gm)

$$= \frac{\text{charge passed (coul)}}{96,500 \text{ (coul per gm-at. wt.)}} \times \frac{\text{at. wt. of material}}{\text{valence of material}}. \quad (21-7)$$

Incidentally, the recurring quantity of 96,500 coulombs is given the convenient name *one faraday* of electric charge, so that we may say simply: in electrolysis, 1 faraday releases 1 gm-atomic-weight of any monovalent material,  $\frac{1}{2}$  gm-atomic-weight of divalent materials, etc., or, in terms of concrete examples, 1.008 gm hydrogen, 31.78 gm copper, 40.59 gm antimony.

**PROBLEM 21-28.** The element barium (Ba, at. wt. 137.36, valence 2) was also discovered by Davy during electrolysis. If you repeat an experiment of this kind and for one hour pass a current of one ampere through molten barium chloride ( $\text{BaCl}_2$ ), how much barium and how much chlorine do you expect to collect? (*Note.* 1 ampere = 1 coulomb/sec.)

**21-16 Atomicity of charge.** Now that the hard-won concepts of atom and electron are so familiar to physical scientists, it is easy to visualize a mechanism that will explain the main phenomena of electrolysis. In such simple cases as the breakdown of molten sodium chloride one can think of the liquid as made of separate atoms of sodium and chlorine. But since Na migrates to the cathode and Cl to the anode, we must account for the influence of electric forces on the atoms; each Na atom after the breakup of the molecule must somehow be positively charged, and each Cl atom must be negatively charged. The passage of electricity through the liquid is then accomplished by the movement of *charged* atoms, or *ions* (from the Greek word for *to wander, to go*).

At this point enters a theoretical picture of the molecule: we are led to the assumption that both Na and Cl, in the pure state of the element sodium and the element chlorine, are ordinary electrically neutral atoms by virtue of a perfect balance between the positive and negative components that make up the atom. Specifically, the negatively charged parts of the atom are the electrons, each having the same definite mass  $m_e$  and charge  $e$ . Now when one neutral atom of Na and one of Cl together form one molecule of NaCl in the solid salt, one of the electrons from the sodium atom will attach itself to the chlorine atom (later we must explain why

this should happen), so that on separation during electrolysis the Na ion lacks one electron and the Cl ion has one electron more than when in its neutral state. These ions can be symbolized by  $\text{Na}^+$  and  $\text{Cl}^-$ . Similarly, a molecule of copper sulphate,  $\text{CuSO}_4$ , splits during electrolytic dissociation into  $\text{Cu}^{++}$  and  $\text{SO}_4^{--}$  ions by the transfer of two electrons.

An explanation of Faraday's law of electrolysis as summarized by Eq. (21-7) suggests itself. Every time one  $\text{Cl}^-$  ion reaches the anode, it gives up its excess electron to that electrode, becomes neutral, and rises eventually in a bubble with other such chlorine atoms. At about the same time, an  $\text{Na}^+$  ion will pick up one of the electrons at the cathode and settle out as neutral sodium metal; we may imagine that this neutralizing electron was made available when the  $\text{Cl}^-$  ion gave up its excess electron to the circuit on the other electrode. Thus for each molecule of  $\text{NaCl}$  electrolyzed into its elements, the charge corresponding to *one* electron moves through the external circuit (wires, battery, and current meters).

This hypothesis leads at once to an estimate of the value of the charge  $e$  on one electron: We know that the passage of 96,500 coulombs of charge results in the electrolysis of one mole or  $6.02 \times 10^{23}$  ( $N$ ) atoms of a monovalent element. At the rate of transfer of one electron per atom, we have

$$96,500 \text{ coul/mole} = (N, \text{ the number of atoms/mole}) \times (e, \text{ charge in coul/electron})$$

$$\text{or} \quad e = 96,500/N, \quad (21-8)$$

from which we can calculate that the charge per electron is  $1.6 \times 10^{-19}$  coul or  $4.8 \times 10^{-10}$  statcoul. The confirmation of this value for  $e$  in several independent and entirely different experiments (e.g., Millikan's oil drop experiment, spectroscopic data) gives us confidence that the whole physical picture leading to this point is plausible. Or conversely, Eq. (21-8) together with separate determination of  $e$  allows us to compute a value for Avogadro's number  $N$ , a value that is indeed in agreement with the results of other experiments.

This concordance of numerical results obtained by widely different methods shows us again the interconnectedness of physical science. From isolated observable facts and limited hypotheses we fashion, as it were, a net or web, and the constants of nature—the constant of universal gravitation, the values of  $N$  and  $e$ , the atomic weights, and the like—are at the points of intersection of different strands, the knots that hold together the whole structure.



The two ideas of importance introduced here are, first, the atomicity of matter and, second, the "atomicity" of electric charges (i.e., charges in liquids are transferred by ions bearing 1,2,3, . . . electrons more or less than when the atoms are neutral). A third idea is also implied, namely, that the electron is ordinarily a part of the atom. But all three postulates were far from evident when Faraday phrased his generalizations on electrolysis in 1833. Atomicity of matter soon became acceptable on other grounds (through chemistry and kinetic theory) but the existence of discrete charges called electrons was first proposed in definite form only in 1874 by G. Johnstone Stoney, and was not completely acceptable for over 20 years more, when precise measurements of the mass and charge on the electron became available through the study of beams of electrons in cathode-ray tubes. The part that electrons play in atoms was then the next major puzzle; its solution, which is the substance of the next chapters, forms one of the most exciting portions in contemporary physical science.

#### Further References

Michael Faraday, *Experimental Researches in Electricity*. London: J. M. Dent and Sons (1914, 1951). Everyman's Library edition. A classic of experimental science. See particularly Part V on electrolysis. Excerpts in W. F. Magie, *op. cit.*, pp. 492-498, W. S. Knickerbocker, *op. cit.*, Chapter XXVI, and J. W. Knedler, *op. cit.*, pp. 447-502.

W. F. Magie, *op. cit.* (Ch. 2), pp. 400-403, 408-420. Excerpts from Franklin and Coulomb. Also W. S. Knickerbocker, *op. cit.*, Chapters X and XXV.

R. A. Millikan, *Electrons (+ and -), Protons, Neutrons, Mesotrons and Cosmic Rays*. Chicago: University of Chicago Press (1947, also earlier editions). The early chapters make excellent reading at this stage.

L. W. Taylor, *op. cit.* (Ch. 5). Chapter 40, historical treatment of electrostatics; Chapter 51 on cathode rays and the properties of the electron.

## CHAPTER 22

### PROPERTIES OF LIGHT

**22-1 Introduction.** The main theme of Part G is, you will recall, the structure of the atom. Chapter 21 dealt with electrostatics and some electrochemistry, and revealed one clue toward the problem of atomic structure—the concept that atoms can contain electrons. Further ideas concerning the detailed structure of atoms came at about the same time from an entirely different field of physics, to which we now turn: the study of light. A brief glance at the historical development of ideas during the last part of the 19th century will set the stage for the introduction of the modern theory of light and its bearing on the evolving picture of the atom. Even so, we shall touch here on only a few topics of the large field of optics.

**22-2 Electromagnetic waves and ether.** We open this inquiry with work done between 1865 and 1873 by the great British scientist James Clerk Maxwell\* (1831-1879) (Fig. 22-1). Its importance for physical science has often been said to approach that of Newton's work, but let us here be content with the very briefest summary. Maxwell's theory of the propagation of light waves is now referred to as the *classical theory*. He first considered what happens when an electric current oscillates along a straight piece of wire, or circulates in a loop of wire. His theoretical conclusion was that electrical energy is lost, that it is radiated away from the electric current and spreads out, wave-like, in all directions. By this is meant that any electric charges in a distant obstacle in the path of the radiation, say in a second wire, will be set into oscillation at the same frequency as that of the original source, the first wire; any magnetic poles present in the obstacle will also undergo similar vibrations, and the obstacle itself will feel a slight mechanical pressure. Technically we can put all this into different language: a vibrating or circulating electric charge (called a transmitter or oscillator) sets up in the region all around it a fluctuating electric and magnetic field; when other electric charges or magnetic poles are introduced into this fluctuating electromagnetic field (as in an antenna or re-

---

\* This is the same man whose mathematical prowess, in 1860, produced the first proof on theoretical grounds that the distribution of molecules in a heated gas should, in the kinetic theory, follow the general shape of the curve in Fig. 20-7.



Yours truly,  
J. Clerk Maxwell

FIG. 22-1. James Clerk Maxwell (1831-1879).

ceiver) they are acted on by electric and magnetic forces that change with the same periodicity as the transmitter's original oscillations. Figure 22-2 shows the electric and magnetic field intensity vectors at one point  $P$  in the electromagnetic field, at eight successive moments. If you concentrate your attention on the behavior of the electric field vector exclusively—it is the only one that we shall pay much attention to—a useful analogy will suggest itself: if some receiver, perhaps a metal wire containing electrons that are free to move, is placed at point  $P$ , the fluctuating electric field will act on these charges in the same manner as a wave passing along a rope will agitate the particles of one section of the rope. In fact, it is perfectly correct to say that an electromagnetic *wave* is passing by this point  $P$  in space, and to refer to the energy responsible for the

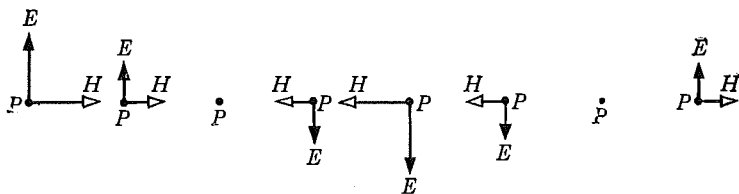


FIG. 22-2. Electric ( $E$ ) and magnetic ( $H$ ) field intensity vectors at the same point  $P$  and at eight consecutive moments.

motion of the charges as electromagnetic energy or electromagnetic radiation.

Maxwell's theory also predicted that metallic objects would reflect a beam of incident electromagnetic energy like a mirror; that on entering other obstacles, such as a sheet of glass, the path of the beams would be bent; and most startlingly, that this radiation must travel through vacuum or air with the speed of about  $3 \times 10^{10}$  cm/sec, a value equal to that for the speed of propagation of light. These and similar predictions of Maxwell's theory all agree at one point: the radiation expected on theoretical grounds from vibrating or rotating currents should be found to behave in every way *like light*. Even the fact that light exerts a pressure was well known from the observation that the material in the tail of comets always points away from the sun at every position along the comet's path (see Section 16-6).

We must therefore investigate the inevitable suggestion that "light itself is an electromagnetic disturbance." At this same point we are also bound to ask for a convenient mental picture of a light wave traveling through space.

In Maxwell's time, the question of how light is propagated seemed to have achieved an uneasy settlement. Newton's thought (not as dogmatically entertained by the master as by his later disciples) had been that light might be mainly explained by the propagation of minute corpuscles through space.\* But the alternative view, presented in rather more definite detail by Huygens and others, had held that light was a minute wave motion in an all-pervading ether, a *mechanical disturbance* like sound, or like some earthquake waves. The English scientist Thomas Young and his younger French colleague, Augustine Fresnel, had fought to put the wave theory on a sound theoretical and experimental basis in the

\* "Are not the Rays of Light very small Bodies emitted from shining Substances? . . ." (*Opticks*)

early part of the 19th century. Finally it had come to a test between the two rival pictures when it was generally agreed that on Newton's model, light *corpuscles* should travel faster in water than in air (the "corpuscles of light" being acted on by accelerating forces of attraction on approaching the surface), whereas light *waves* should travel more slowly in water. In 1850, Foucault showed by experiment that the latter was indeed the case, and the wave theory seemed firmly established.

The question then remained, What is the ether, the medium in which light waves can progress even when all air or other ponderable substances are removed? Magnetic and electric forces, perhaps also gravitational ones, all demanded overlapping ethers. There had already been some progress toward a broadening of the conception of light. Sir William Herschel had discovered in 1800 that there always existed in the spectrum of hot solids, along with visible light, some radiation of longer wavelengths (beyond the red, namely the *infrared*). The suggestion that this "invisible light" was identical with radiant heat was proved when heat rays were shown to exhibit all the characteristics of light—refraction, interference, polarization, and so forth. Already there is a simplification here: Newton had thought that radiant heat might need a separate ether for its propagation, now the light-bearing or *luminiferous* ether could serve a double purpose.

But light still made extravagant requirements on its medium of propagation. We remember again that the wave theory of light interpreted light as a *mechanical* disturbance in ether. But experiments on the speed of all ordinary mechanical disturbances in solids, liquids, and gases had proved that the speed of such waves is proportional to  $\sqrt{\epsilon/\rho}$ , where  $\epsilon$  is a measure of the elasticity of the medium and  $\rho$  its density. For speed as great as that of light the medium of propagation, the ether, would have to show simultaneously a very high elasticity and an extremely low density. The low value of the density by itself did not create conceptual problems; a high degree of rarefaction of the ether was expected, for this would account for the well-known facts that the planets are not noticeably impeded in their motions through the ether, and that all attempts to separate and identify the ether by mechanical means had consistently failed. This point had been well expressed long ago by Newton in the *Opticks*:

“. . . And therefore to make way for the regular and lasting Motions of the Planets and Comets, it's necessary to empty the

Heavens of all Matter, except perhaps some very thin Vapours, Steams, or Effluvia, arising from the Atmospheres of the Earth, Planets, and Comets, and from such an exceedingly rare Aethereal Medium as we described above. A dense Fluid can be of no use for explaining the Phaenomena of Nature, the Motions of the Planets and Comets being better explain'd without it. . . . And if it be rejected, the Hypotheses that Light consists in Pression or Motion, propagated through such a [dense] Medium, are rejected with it."

But even if we assign to the density of ether the lowest plausible values, the great speed of light would still demand our now very diluted medium to possess the elastic properties of steel. Scientists of today, who more than their ancestors have grown up without feeling the need to account for phenomena by mechanical models, might find it more difficult to conceive of a medium with such contradictory properties than to think of light propagation through space without the benefit of an intermediate material medium. But it should be clear that in the 19th century, instead of discouraging the proponents of ether models, it increased their efforts. It seemed that the mind could not contemplate waves without visualizing a specific medium and, in fact, Maxwell's own thinking and his derivations absolutely depended on his assumption of certain rather complex mechanical models for the ether. To this day there are echoes of this conceptual struggle, which we recognize as having its equivalent in physics since the time of Aristotle.

The merger of the two ethers for radiant heat and for light into *one* had been the first step in the direction of unification, of making one ether do the job of several. Maxwell's work was a second, big step; now the propagation medium for magnetic and electric effects took over the function of transmitting light and heat waves also. In the long run, however, this simplification had a very unexpected effect: when the time came to abandon the ether concept, it was easier to do so by giving up one general ether than it would have been to disprove and dismiss several ether concepts separately.

That time came in 1905. The one feature all hopeful ether models had in common was that they provided a medium for the propagation of light which existed apart from the source and the observer, just as air is a medium for sound transmission. To pursue this analogy further, consider how one measures the speed of sound: one might station a loudspeaker and a microphone many meters

apart and time the progress of a short pulse of sound along that distance. At a temperature of  $0^{\circ}\text{C}$ , the speed in air is ordinarily 330 m/sec. But if this experiment were made on an open railway flat-car, or in some other "laboratory" that moves with respect to the medium (air), then the speed measurement for sound would be affected by our own motion. For example, if the sound is directed to a microphone at the rear end of the flatcar, it will get there sooner if the car moves forward than if the car stands still. Similarly, the measured velocity of light should vary depending on the motion of the "laboratory" through the ether and the placement of the equipment relative to the motion. Our earth itself is a fast-moving "platform" as it revolves about the sun and rotates on its own axis; light moving along a meter stick that points to the East is caught up in the relative eastward motion of the ether through which the earth moves, and arrives at the other end sooner than if it had to travel the opposite way or along a North-South line.

This difference in the speed of light, which should give the velocity of the earth through the ether or, equally well, the relative velocity of the ether past the earth, was essentially what the American experimenters A. A. Michelson and E. W. Morley attempted to find in 1887. On the basis of their technique one could have expected a clear and convincing positive result—and yet, no difference of speeds was observed! Here was a new Postulate of Impotency: One cannot measure the relative velocity of the ether.\* After a period of adjustment and speculation, with varying degrees of success, this famous experimental result led to a more far-reaching postulate, one of the basic pillars of Einstein's Relativity Theory: The measured velocity of light in space is constant no matter what the motion of the observer or of the source may be. On the one hand, this extends the previous discussion on the Relativity Principle (Section 3-5) to include optical experiments in the statement "All laws of physics observed in one coordinate system are equally valid in any other coordinate system moving with a constant velocity relative to the first." And on the other hand, the acceptance of the postulate of constancy of light as a fundamental principle relieves us of the need to account for the propagation of light by an ether model. For if the principle is right, as would be found by checking

---

\* That the earth was not simply "dragging" the ether in its vicinity was made clear in other experiments.

experimentally all consequences derivable from it, then we are better off to accept the principle as a basic point of departure than we should be if we continued the dubious quest for an ether free from all internal contradictions and in accord with all experiments. The chief point is that one must allow that a useful principle, even one unfruitful of convenient mental images, is to be preferred over a clear model which is not consistent logically and not in accord with empirical facts.

This point reminds us of the fate of the caloric, the medium whose function was to provide a visualizable mechanical model for the processes of heat transfer. It was challenged by Rumford and Davy on the grounds that it failed to exhibit the behavior expected of material media—specifically, that it had no measurable weight, and could, in friction, be created without limit and without drawing from the materials out of which it arose. The comparison between the fate of the caloric and of the ether is, however, faulty in one detail. The caloric was eventually replaced by a scheme built on another visualizable model, the kinetic theory picture of the atom, whereas the ether was replaced by a principle, and by the habit of speaking of “the field” as the arena for physical processes that occur in *vacuo*, in the absence of ponderable matter. If we wish, we may conceive of light for the present as progressing through space by the propagation of electric (and magnetic) field vectors (see Fig. 22-3 for a “snapshot” of a region through which light travels, with some field vectors drawn in). The distance between successive instantaneous maxima is one wavelength  $\lambda$ , and the frequency of the wave  $\nu$  (Greek letter *nu*) is the number of crests passing a given point per second. The velocity  $c$  is given by the usual relation  $c = \lambda\nu$ . Our elimination of the ether does not, at this point, invalidate the rest of the Maxwellian theory of electromagnetic waves.

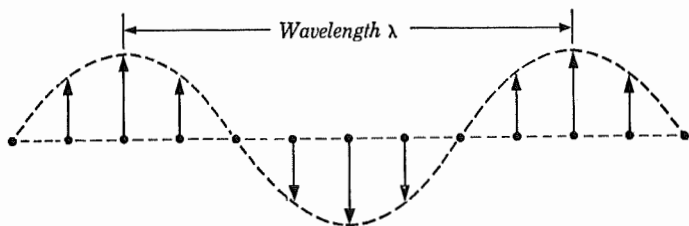


FIG. 22-3. Propagation of electromagnetic wave through space; only electric field vectors are indicated at 13 equally spaced points.





A handwritten signature in cursive script that reads "H. Hertz".

FIG. 22-4. Heinrich Hertz (1857-1894).

**22-3 Hertz's experiments.** If Maxwell's theory of light as an electromagnetic radiation emitted by the periodic fluctuation of currents is correct, a heated object, e.g., the metal filament of an incandescent lamp, sends out light by virtue of the agitation on the filament of submicroscopic electric charges such as one expects to be distributed throughout neutral bodies in balancing amounts.

We cannot decide at this point whether these charges are parts of the atoms in the emitter or separate entities such as electrons or positive particles, so permit us to call them for the moment, in a general way, "oscillators." Somehow these oscillators are able to emit electromagnetic energy that our eye interprets as light, the frequency  $\nu$  of the light waves being equal to the periodic frequency  $f$  of the oscillators.

This "classical" interpretation of light emission might be more acceptable to us if we could refer to direct experimental confirmation. One very obvious test might be to produce an oscillating

current in a wire at a frequency equal to that of, say, green light, and watch whether green light is actually released into the surrounding region. Even before the days of radio and microwaves it was easy enough to produce electric oscillations in wires, but the frequencies achieved were relatively low in Maxwell's day, and even today the higher frequencies of electric oscillation in the most modern equipment can hardly be pushed above  $10^{12}$  sec<sup>-1</sup>, far below the frequency of visible light ( $\sim 10^{15}$  sec<sup>-1</sup>). Consequently, a direct test seemed out of the question, and partly for this reason Maxwell's work did not attract a great deal of attention for about 20 years.

There is, however, another point of attack. The gifted German physicist Heinrich Hertz (1857-1894) (Fig. 22-4) undertook in 1887-1888 to show that at least the main premise of Maxwell's view was correct, namely, that an oscillating electric current (he attained a frequency of about  $10^9$  sec<sup>-1</sup>) does indeed send out electromagnetic waves that have all the characteristics of light except, of course, visibility. Hertz's proof succeeded so brilliantly that hardly one important physicist remained unconvinced.

In essence, Hertz's experimental method was simple. Two rods, each about 5 inches long and terminated at one end by a small polished metal sphere, were fixed along a line so that a small air gap separated the spheres (Fig. 22-5).

A pair of wires led from the spheres to a device supplying short pulses of very high potential difference, and during those intervals brought to the two spheres large amounts of electric charge of opposite signs, until the



FIG. 22-5. Spark gap in Hertz's oscillator.

point was reached where a spark jumps across. The air in the gap remains conducting for a short time and so provides a path for these charges as they oscillate from one rod to the other until electrical equilibrium is achieved. Then, the air being restored to nonconductivity, the stage is set for the next "voltage" pulse and another spell of oscillations by the same processes.\*

If Maxwell was right, this arrangement should, during each spark, give off electromagnetic waves of the same frequency as the

---

\* Let us take notice of a short and somewhat mystifying remark of Hertz's which will assume strange importance later, namely, that "it is essential that the pole surfaces of the spark gap should be frequently polished" to assure proper operation of the spark.

oscillations, up to about  $5 \times 10^8 \text{ sec}^{-1}$  by Hertz's estimate, corresponding to a wavelength of about 60 cm. Furthermore, this radiation should spread out from the rods and be detectable by its production of a fluctuating current in a wire some distance away, as has been pointed out.

Hertz's first triumph was to observe this very effect, even when he placed the receiving wire (antenna) many yards from the sending oscillator. This new invisible electromagnetic radiation, still sometimes called Hertzian waves, is of course identical with our now familiar radio waves. Hertz showed next that the radiation could be made to go through all the characteristic behavior of light: reflection, focusing into a beam, refraction, interference, polarization, and so forth. One interesting test, the direct determination of the velocity of this radiation, was beyond the instrumentation available to Hertz. Experimentation conducted after 1895 confirmed the assumption that the speed was the same as that of light.\*

One conclusion was soon almost universally accepted: since Hertzian waves and light waves behave so much alike, light too must be caused by the rapid oscillatory motion of charged particles, perhaps those in the atoms of the light emitter. Here then was the model for light emission. And in terms of this picture, a great many details became clear and could be integrated into one scheme—the mechanism for transmission of light through crystals and other (transparent) objects, the reflection from metal surfaces, and so on. One particularly impressive, even sensational, confirmation of the theory was provided by P. Zeeman, who discovered (1896) that a glowing gas sends out light at changed frequencies if the emitting material is placed in a strong magnetic field. This effect could be explained for many cases by a straightforward adoption of the fundamental Maxwellian idea that light energy is released by

---

\* The demonstration of Hertzian waves proved to be not only the intended triumph of Maxwell's theory but also the unintended beginning of radio. It is perhaps not amiss to contemplate briefly that strange fruit on the tree of science. Consider an ordinary radio set from all the following points of view—symbolically, as it were: how the first hint of radio waves originated in the mathematical speculations of Maxwell, and through the ingenious experiments of Hertz came to the notice of the young Italian Marconi. Consider how even these three men, in the continuity of their work and the differences of their nationalities, religion, and training, exemplify the interdependence of man, the basis of the humanistic ideal. Then consider the meaning of radio as a huge industry in the economics of the nation, and how society rarely uses this tool for mass education, often for mass propaganda.

the vibration of charged particles, since the observed change in light frequency brought about by the field is equal to that calculated from Maxwell's theory.

**22-4 Continuous emission spectra.** Our newly gained, though as yet rather vague, model of the emitter of light, the submicroscopic electric oscillator, raises our hopes for the solution of some long-standing puzzles about radiation: Can we not now begin to *explain the colors (frequencies) and distribution of light\* energies* sent out by different sources of light?

Before we can grapple with this problem we must turn to a brief account of the large amount of experimental knowledge available by the turn of the 20th century. Newton had shown, in the "celebrated experiment," that light from the sun could be resolved by a prism into a spectrum (Newton's word) of various colors; and just such a "rainbow" spectrum had since been obtained by examining the light from *all* glowing solids (filament, carbon in arc) and glowing liquids (molten metals). With only a little idealization, these emission spectra all have three features in common:

(a) They are *continuous* spectra, that is, there are no gaps, no "color" bands missing from the spectrum, even if measurements of the radiation are carried into the infrared or ultraviolet region.

(b) All glowing solids and liquids, *no matter what their chemical nature*, send out light with about the same color balance if they are at the same temperature. A red-glowing coal has that color not because there are no other colors but red being radiated from it, but because of all the visible radiation that in the "red" region is most intense, and an iron poker put into the coals, after reaching the same temperature, glows with the same light.

This must be put more precisely. The wavelength of each color (and, by means of the relation  $c = \lambda\nu$ , the frequency also) can be measured with great accuracy by means of a spectroscope.\*\* Approximate ranges in wavelength for different colors are given below. Note that wavelengths are usually not given in centimeters but in units of  $10^{-8}$  cm, called 1 *angstrom* or 1 Å, in honor of the Swedish spectroscopist A. J. Ångström.

---

\* The identity of the character of ultraviolet rays, visible light, infrared (heat) rays, and longer wavelength (Hertzian) waves having been established, we shall from now on refer to all these as "light."

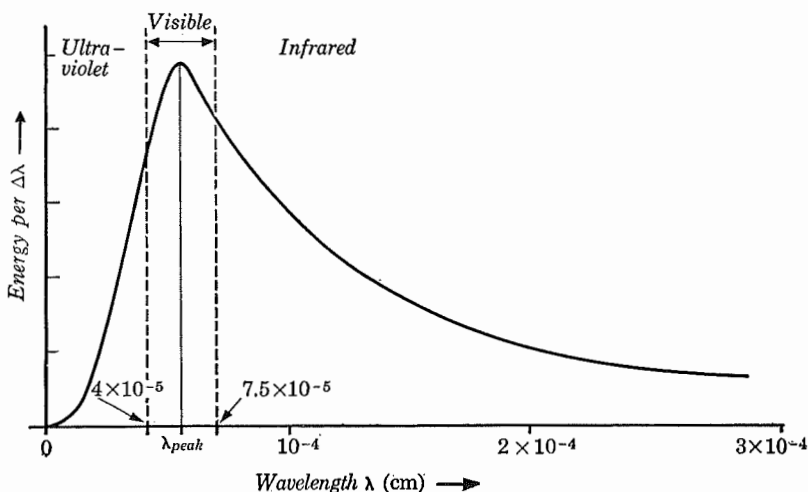
\*\* Although not a prism but a diffraction grating is used for the determination of absolute wavelengths. See bibliography for detailed discussion.

Table 22-1. Wavelengths of Colors in the Spectrum

<i>Color</i>	<i>Range of <math>\lambda</math> in <math>\text{\AA}</math> (<math>10^{-8}</math> cm)</i>
Red	6100 to about 7500
Orange	5900 to 6100
Yellow	5700 to 5900
Green	5000 to 5700
Blue	4500 to 5000
Violet	about 4000 to 4500
(Infrared)	Longer than 7500
(Ultraviolet)	Shorter than 4000

The exact amount of energy contributed by the components of different wavelengths in a beam of light can now also be measured with precision, although originally the details of experimentation were extremely delicate and difficult. For the infrared ranges, the energy can be detected conveniently by sensitive heat-measuring devices (thermopiles, bolometers), and in the visible and ultraviolet ranges by photographic exposures (so used since about 1840), photocells, and the like.

A typical distribution curve of radiated energy *vs.* wavelength, specifically for a temperature corresponding to that on the surface of the sun ( $6000^{\circ}\text{K}$ ) is seen in Fig. 22-6. In this case there is rela-

FIG. 22-6. Distribution of energy in light emitted from a glowing body at  $6000^{\circ}\text{K}$ .

tively little energy at very short wavelengths. The peak of the curve falls into the green part of the visible region, and the largest share of the total energy emitted is in the infrared; our eye perceives only the particular mixture of "visible" radiation, and compounds it to give the impression of white light. The chief points of interest here are that such a curve can be drawn for every temperature of the emitter and that all emitters of the same temperature, regardless of chemical composition, yield about the same curve.\*

(c) The third feature which all hot solid or liquid emitters have in common is that their energy distribution curve shifts with changing temperature  $T$ , as shown in Fig. 22-7 for four temperatures. With increasing  $T$ , more energy is radiated in each wavelength region, and the peak in the curve moves toward shorter wavelengths.

The exact data were summarized by the German physicist Wilhelm Wien\*\* in 1893 in the so-called "Displacement Law," which in part states that for a perfect emitter of continuous spectra, the product of  $\lambda_{\text{peak}}$  (in cm) and its temperature (in  $^{\circ}\text{K}$ ) is a constant with the empirical value of  $0.2897 \text{ cm}^{\circ}\text{K}$ :

$$\lambda_{\text{peak}} (\text{cm}) \times T (^{\circ}\text{K}) \doteq 0.29 \text{ cm}^{\circ}\text{K}. \quad (22-1)$$

For example,  $\lambda_{\text{peak}}$  for the sun is, by experiment, about  $5.5 \times 10^{-5} \text{ cm}$ . It follows that  $T = 5300^{\circ}\text{K}$ . But this figure is a little too low, partly because some of the energy radiated by the sun, particularly at short wavelengths, is absorbed in the atmosphere before reaching

\* To be more accurate, the radiation has the distribution shown for emitters that, when cool, are perfect absorbers of light. These are called "blackbodies," but we can here disregard deviations from blackbody radiation.

\*\* Wien received a Nobel prize in 1911 for his work on the theory of radiation.

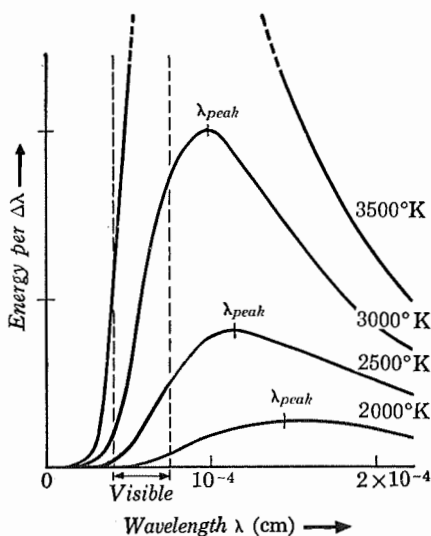


FIG. 22-7. Distribution of energy in light radiated from a glowing solid at different temperatures.

our instruments, so that the true  $\lambda_{\text{peak}}$  is lower; and also because the sun is not the ideal radiator (so-called "blackbody radiator") for which Eq. (22-1) is designed. Allowing for these difficulties, one is led to a value of about  $6000^{\circ}\text{K}$  for the sun's surface.

Provided Wien's displacement law holds even in extreme temperature ranges, we may at once make some surprising calculations: The radiation from distant stars can be analyzed to find where  $\lambda_{\text{peak}}$  falls. From this we may calculate approximate stellar surface temperatures. For the "hot" stars, whose  $\lambda_{\text{peak}}$  is small and lies in the blue end of the spectrum or even beyond, the temperatures are correspondingly higher than that of the sun (e.g.,  $14,000^{\circ}\text{K}$  for Vega). The cooler, reddish stars have larger values for  $\lambda_{\text{peak}}$ , and the calculated values of  $T$  are consequently lower (e.g.,  $2500^{\circ}\text{K}$  for Antares).

PROBLEM 22-1. What is the wavelength of the most intense radiation for Vega, for Antares, and for a body at room temperature? Are there qualifications to your answer?

PROBLEM 22-2. The total area under the curve for intensity of radiation *vs.* wavelength is proportional to the total energy radiated throughout the spectrum. Trace the curves in Fig. 22-7 and find an approximate law connecting the total energy radiated and the temperature of the emitter. (In the ideal case the relation is total energy  $\propto T^4$ .)

PROBLEM 22-3. On the basis of our model of light emitters, would you expect to find a distribution curve such as shown in Fig. 22-6 or instead that all energy is sent out at *one* frequency? Should chemical and physical consistency influence the light emitted? What might temperature have to do with the distribution of intensity *vs.* wavelength? (Defend your conclusions, but be frank about vague guesses.)

The displacement law could actually be derived theoretically from the classical (Maxwellian) theory of light emission, and that was indeed another impressive victory! But at that very same point we encounter serious trouble. Surely we should also be able to explain the exact shape of the characteristic emission curve as a whole by derivation from the basic model of emission. This would involve first a calculation of the fraction of individual oscillators for each range of emitted frequencies, then the addition of their individual contributions to find the over-all pattern.

*But all such attempts proved fruitless.* Specifically, Fig. 22-8 compares the actual emission curve with the prediction made by two modifications of the classical theory that differed from each other by some of the assumptions introduced in the classical model.

Neither modification yields the correct curve. We recall instances of similar dilemmas in the history of science. The most memorable case is perhaps Kepler's attempt to derive from the Copernican theory a path for the planet Mars that would fit the observations. He failed, just as we now find ourselves unable to fit theory and fact. His solution was not to forget about the fairly small discrepancy—

that is permissible only if one can account for it exactly by computation of the secondary effects at work, or at least tentatively by a consistent plausibility argument, as Galileo accounted for small deviations from the law of free fall by citing the effects of air friction. In Kepler's case, the small discrepancy provided a real challenge, the point of departure for a profound change of outlook in the whole field. You will recall that Kepler turned from the unfulfilled theoretical prediction to find a mathematical formulation of the actual (elliptical) planetary path, and to adopt that as the starting point of a modification of the Copernican heliocentric theory. This led him next to the great discovery of the three mathematical laws of planetary motion. Finally, a good physical interpretation of Kepler's laws was given by Newton in terms of the gravitational forces, thereby "explaining" Kepler's empirical findings.

A very similar turn of events will soon be revealed in our coming discussion of the emission problem: we shall see that the obstinate mismatch of the curves in Fig. 22-8 prompted first a mathematical formulation of the actually observed curve; then came a physical interpretation of the mathematical formula which modified the old theory of light emission as drastically as Newton's work improved on the Copernican concept of the universe. For the moment, however, we must look at greater length at more of the experimental results that sparked the revolution in physical science, the quantum theory.

**22-5 Line emission spectra.** The foregoing section concerned only the spectra from glowing solids or liquids, and incidentally also

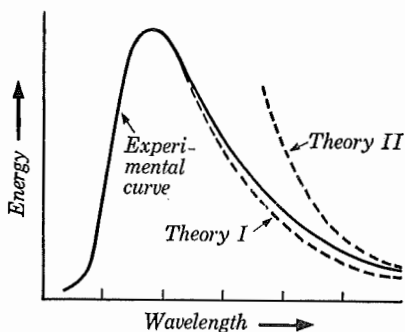


FIG. 22-8. Comparison of experimental emission curve with predictions from two classically based theories.



from gases at those extreme conditions of density and temperature found in our sun and other stars but not in our laboratories. We noted that these continuous emission spectra aid in finding source temperatures (from  $\lambda_{\text{peak}}$ ), but give no clue about the chemical composition. Now we turn to another type of spectrum.

It had long been known that light is emitted by gases and vapors when "excited" by the passage of an electric spark or arc or, as in the case of neon lights, by a continuous electric current established through a thin tube filled with some of the gaseous material (Fig.

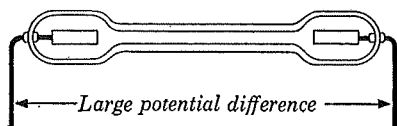


FIG. 22-9. Electric discharge tube. Current from one sealed-in electrode to the other causes the gas to emit light.

22-9), or when volatile matter is put into a nonluminous flame. Furthermore, the light so emitted, when resolved into its components by the prism spectroscope, was found to have a spectrum startlingly and fundamentally different from the continuous emission

spectra of glowing solids and liquids: gases and vapors have *line emission spectra*, that is, instead of a solid band of changing "rainbow" colors they show light only at some well-defined wavelengths, with dark gaps between neighbors. As seen through the spectroscope or photographed on a plate in a spectrograph, the spectrum therefore appears as a set of lines, usually irregularly spaced, some very bright, others less so.

The second point of difference with respect to continuous emission spectra is that the line emission spectra are markedly different for different radiating elements. Each substance has its own characteristic pattern of wavelengths throughout the whole observable region (Fig. 22-10). The unaided eye synthesizes these separate lines and recognizes the mixture of colors as reddish for neon, pale blue for nitrogen, etc. Some materials reveal a most complex emission spectrum, others

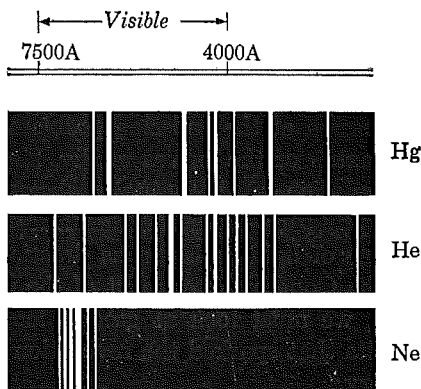


FIG. 22-10. Sections of the line emission spectra of mercury, helium, and neon. (Redrawn from photographic record.)

are far simpler; iron vapor shows about six thousand bright lines, whereas sodium has only two strong, close yellow lines in the visible range. The great variety of patterns and the individual spacing of lines seemed quite unaccountable; why should gases emit line spectra at all, and why should closely related elements present such differing line patterns? All this was destined to remain most puzzling for nearly three generations. But at any rate, each material could be identified from its unique line emission spectrum as surely as criminals can be traced by means of their characteristic fingerprints.

Rapidly, a whole art of spectroscopy arose; more and more powerful prism and grating spectroscopes became available to chart the exact wavelengths of these patterns. As early as the 1830's, the suggestion was made that the presence, identity, and abundance of materials in ores, etc., might be determined from the type and intensity of the emission spectra. And in fact, the physicist G. R. Kirchhoff and the chemist R. W. Bunsen at the University of Heidelberg, who were the foremost pioneers in the development of spectroscopy, discovered two elements, rubidium and cesium, in the vapor of a mineral water by noting hitherto uncharted emission lines. It was the first of a brilliant series of such discoveries, and the origin of a technique that has given us the speedy chemical analysis of small samples by routine spectroscopy, a vital tool in today's research laboratory and industry alike. Another spectacular achievement that caught the imagination of all Europe was the spectroscopic analysis of meteorite material. By comparing the emission spectra of vaporized meteorites it was established that these contained only well-known elements present on our earth, a direct confirmation of the unity of terrestrial and celestial science that would have gladdened the hearts of Galileo and Newton.

**22-6 Absorption line spectra.** In 1814, the German optician Joseph Fraunhofer studied a peculiarity in the appearance of the continuous spectrum of sunlight. When light was passed through a narrow slit and its visible spectrum very carefully examined with

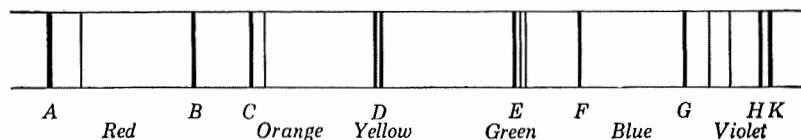


FIG. 22-11. Fraunhofer lines in the sun's continuous spectrum; only some of the most prominent lines are shown.

good prism systems, the continuity of colors appeared to be disrupted by a series of fine, steady, irregularly spaced *dark lines* (Fig. 22-11). Fraunhofer counted more than 700 such lines (now, with better instruments, we know of over 15,000) and assigned the letters *A, B, C, . . .* to the most prominent ones. He also observed many such lines in the spectra of some bright stars.\*

The key observation toward a better understanding of both the absorption and the emission spectra came in 1859. By that time it was generally known that the emission line from the heated vapor of sodium metal had the same wavelength as the very prominent dark line\*\* in the solar spectrum to which Fraunhofer had assigned the letter *D*.

What Kirchoff now found looks simple in retrospect. In essence the argument is this: When the light from a pure, glowing solid is ordinarily analyzed, it does not show dark lines crossing the continuous emission spectrum, but if (as shown in Fig. 22-12) that same light on its way to the prism is now allowed to pass through a suitably large volume of gas or a cloud of vapor, e.g., vaporized sodium metal, then the spectrum does exhibit one or more dark lines of exactly the same wavelength and appearance as do those in the Fraunhofer observation. Each vapor was found to have a characteristic line absorption pattern (Fig. 22-13); evidently each strongly absorbs certain characteristic frequencies from the passing "white" light. And, most interesting of all, the wavelengths of the *absorption* lines in each case correspond exactly to some lines in the catalogue of *emission* patterns for the particular gaseous substance.

We may therefore attempt a preliminary interpretation of the dark lines in solar and other stellar spectra: the missing lines correspond to light that was withdrawn or somehow scattered from the

---

\* It is significant that Fraunhofer's interest in these dark lines was initially a practical one; they identified and labeled portions of the solar spectrum, and so provided reference points useful for comparing the degree of dispersion of colors attainable with different types of glass. Here, as so often in the history of science, the skills of the craftsman and the needs of technology eventually bring to the notice of scientists a new natural phenomenon and techniques for dealing with it; and from such beginnings, in the hands of a gifted scientist, profound and unexpected contributions may follow.

\*\* There are two lines, with wavelengths of 5889.953 Å and 5895.923 Å, but only modern instruments of great precision can resolve them that well, and so we shall speak of *a* sodium line at about 5890 Å.

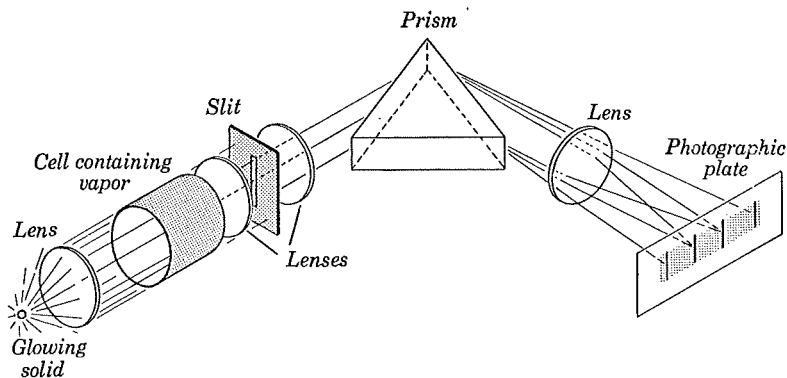


FIG. 22-12. Diagram of spectrometer set up for recording an absorption spectrum.

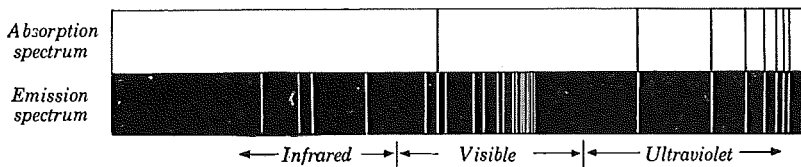


FIG. 22-13. Absorption spectrum of sodium compared with the line emission spectrum of same substance.

total beam on passage through the (relatively cooler) atmosphere surrounding the star and, to some extent, through the atmosphere of our earth. This view leads us to regard the Fraunhofer lines as evidences of the presence of specific elements in the atmosphere of a star. We see at once the enormous potentialities of this theory for astrophysical research.

**PROBLEM 22-4.** How would you decide, on an experimental basis, which absorption lines in the solar spectrum are due to absorption near the sun's atmosphere rather than to absorption by gases and vapors in our own atmosphere?

**PROBLEM 22-5.** How might one decide from spectroscopic observations whether the planets shine by their own light or by reflected light from the sun?

Among the consequences following from this explanation of Fraunhofer lines, three are of particular interest to us now:

(a) Each single Fraunhofer line ought to be assignable to the action of some specific gas. Therefore, when with better techniques a new line was discerned in the green part of the solar spectrum

which corresponded to none of the lines known to be emitted or absorbed by terrestrial material, it was plausible to ascribe it to the action of a gas in the atmosphere of the sun not yet discovered on earth. The appropriate name *helium* was chosen and, indeed, many years later a gas with such an absorption line was found, first as a constituent of many minerals, later also in natural gases and in our own atmosphere.

(b) After absorbing its characteristic colors from white light, a gas may re-radiate it, and generally at the same frequency. This conclusion was original with Kirchhoff. We may understand the essence of his argument quickly in terms of an analogy: A beam of sound containing many frequencies is sent past a mounted and silent tuning fork tuned to 440 cycles/sec. Any sound of that particular frequency will not simply pass by the tuning fork as do all the other frequencies, instead, being "in resonance" with the instrument, its energy will be absorbed as it sets the prongs vibrating. A distant listener along a line between sound source and tuning fork might therefore notice that this one component reaches him with less than the full energy. However, the tuning fork now vibrates, and sends out again the sound energy it received—but in all directions, not just in the direction to which the absorbed sound was traveling. The proportion of sound energy re-radiated *toward* the observer is very small compared with the total energy absorbed. The conclusions from the analogy are, first of all, that a dark line in an absorption spectrum appears dark only because of the relatively much greater intensity of the "background" of unaffected radiation at either side; *some* light is always also re-radiated in the original direction of propagation. Secondly, we expect to see some light re-radiated toward us from a volume of gas while it is absorbing light, even if we view it at  $90^\circ$  to the direction of propagation of the original radiation; for example, we expect from the edges of the absorbing layer of the sun's atmosphere a bright line spectrum corresponding exactly to a *reversed* Fraunhofer spectrum. This should happen during those few moments while the main disk of the sun itself is just hidden in a total eclipse, and this deduction was first and spectacularly fulfilled during an observation of this kind in 1870.

(c) It is not difficult to visualize the basic mechanism for this *resonance* interpretation of absorption in terms of the classical theory of light as it developed soon after Kirchhoff's work. We recall that an "oscillator" corresponds to an electric charge vibrating or revolving at the same frequency as the light radiation thereby emitted.

When such an oscillator is inactive because no energy is being supplied to it (by heating or by an electric discharge), it nevertheless remains "tuned" to its particular frequency, ready to receive energy from a passing light ray of that same frequency, and once in full vibration, it must give up (re-radiate) energy as fast as it is supplied.

These points, which followed previously from the tuning fork analogy, are in general accord with the Maxwellian scheme. But if the classical theory could furnish no explanation for the distribution of intensities in the relatively simple and uniform case of continuous emission spectra, the explanation of bright line emission spectra seemed infinitely more hopeless. Here are some of the conceptual difficulties:

One might argue that the atoms of different elements may well have characteristic structures of their own, so that each elementary atom by itself would emit at the frequency of its particular oscillatory vibration. But the spectra from solids or liquids or very hot and dense gases do *not* show any differences from one element to another, and this can only mean that in those physical states the nearness and incessant mutual collision among atoms blurs the individual vibration characteristics, just as pendulums or springs can no longer vibrate with their free frequencies when connected together in large groups. This picture qualitatively accounts for the difference between the continuous and the line emission spectra, but there is this serious flaw: even in the rarefied gas in a glowing discharge tube, atoms are continually colliding billions of times per second, according to our kinetic theory model of gases, and usually with enormous speeds; yet we do get from them definite line spectra. What could account for the complete blurring of spectra in one case and the lack of interaction among colliding "oscillators" in the other case?

Quite apart from this last difficulty, how are we to account for the exact line pattern itself? In some of the more involved spectra, one might have to visualize an internal structure of "oscillators" in the atom which would make a grand piano exceedingly simple by comparison (to use an analogy then current). Equally disappointing was that the line spectra, while definite for each element, still were apparently without order. They showed no progressive change from one element to the next; as noted, even chemically related elements often exhibited strikingly different arrangements of lines.

A further complexity was introduced by the observation that the

number of observed lines in every case depended on the violence of the electric discharge; further, that while every absorption line indeed corresponded to some emission line in each gas, yet there were always many more of the latter which had no equivalents in absorption spectra. This point conflicted with the idea of resonance radiation, for surely a musical instrument, to return to our analogy, should freely absorb every sound to which it is "tuned" and which, on being played, it can emit.

Another most serious difficulty was that each emission line corresponds to a definite state of vibration for the atomic oscillator. Using the terminology of Chapter 20, we must say that each emission line reveals a separate *degree of freedom* of the oscillator. The principle of equipartition of energy among all active degrees of freedom leads us to expect that at a given temperature  $T$  the oscillator possesses  $\frac{1}{2} kT$  units of energy per degree of freedom. For  $\delta$  spectral lines, and  $\delta$  degrees of freedom, this means  $\frac{1}{2} \delta kT$  ergs per atom, and  $\frac{1}{2} \delta RT$  ergs per mole of the glowing substance. Ordinarily the number of spectral lines ( $\delta$ ) is very large, but on the other hand, the experimental values of specific heats for gases showed clearly that the total energy per mole is only of the order of  $\frac{3}{2} kT$  to about  $\frac{5}{2} kT$ , depending on molecular structure. This contradiction challenges the whole concept of equipartition.

Because the field of experimental spectroscopy after the work of Kirchhoff and Bunsen grew so vigorously, the avalanche of data tended to point up the chaotic features with more and more discouraging certainty. True, there were a few hints of order, by themselves also unexplainable: the general appearance of the spectra of alkali metals showed *some* features of similarity when the regions beyond the visible were included; the wavelength difference between neighboring lines for some parts of the sodium and the zinc spectrum was constant, and groupings of 2 or 3 lines tended to recur in some other spectra; but not much else. This was a period of almost obsessed searching for some hint of a numerical relation among the tantalizing lines, for some mathematical key to decode the message that surely had to be behind the appearances. It is as if we watched again the pupils of Plato trying to reduce the offensive disorderliness of planetary motion to a series of circular motions, or Kepler poring endlessly over his data to find relationship between the radii of planetary orbits and the times of revolution. And, most prominently, we recall the chemists Döbereiner to Mendeléeff trying to find order among the list of chemical elements. We are here very close to a vital process in the evolution of the sciences.

**22-7 Balmer's formula.** As sometimes happens, the first important break in this problem came from a fairly obscure corner. In 1885 a Swiss school teacher, Johann Jakob Balmer (1825-1898) published a paper entitled simply "Notice concerning the spectral lines of hydrogen." His starting point was Ångström's fine measurements of the wavelengths of the four quite well-known visible hydrogen emission lines, commonly named  $H_\alpha$  (red),  $H_\beta$  (green),  $H_\gamma$  (blue),  $H_\delta$  (violet). Quite probably by a straight cut-and-try method, Balmer hit upon a formula that would relate these four wavelengths and which we may write as

$$\lambda(\text{cm}) = C \times \left( \frac{n^2}{n^2 - 2^2} \right), \quad (22-2)$$

where  $C$  is a constant quantity, equal to  $3645.6 \times 10^{-8}$  cm, and  $n$  is a whole number. Specifically,  $n = 3$  if  $\lambda$  for  $H_\alpha$  is to be found,  $n = 4$  for  $H_\beta$ ,  $n = 5$  for  $H_\gamma$ , and  $n = 6$  for  $H_\delta$ . Table 22-2 shows the excellent agreement (to within 0.02%) between Balmer's empirical formula and Ångström's measurements on which it was based.\*

**Table 22-2. Data from Balmer's Paper**

Name of line	Wavelength (Å)	
	Experimental	From Balmer's formula
$H_\alpha$	6562.10	6562.08
$H_\beta$	4860.74	4860.8
$H_\gamma$	4340.1	4340.0
$H_\delta$	4101.2	4101.3

Balmer speculated next what the wavelength of a fifth line might be if it existed. Using  $n = 7$  he finds  $\lambda$  of  $H_\epsilon = 3969.65$  Å, from Eq. (22-2) (check this), and he writes "I knew nothing of such a fifth line, which must lie within the visible part of the spectrum . . . and I was compelled to assume that the temperature relations [in the glowing gas] were not favorable to the development of this line or that the formula was not generally applicable." But then, he continues, a colleague mentioned that more lines had been found in the violet and ultraviolet, in the analysis both of emission from

\* With the sublime self-confidence of a numerologist Balmer comments at this point: "A striking evidence for the great scientific skill and care with which Ångström must have gone to work."



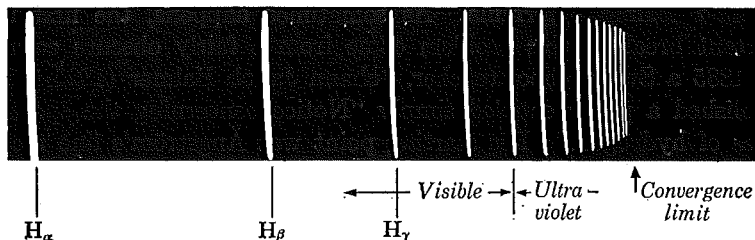


FIG. 22-14. Balmer series of hydrogen emission spectrum.

discharge tubes and of stellar (absorption) spectra. The fifth line did indeed exist, and had the predicted wavelength! In a postscript Balmer relates that just then he received the measured wavelengths for nine more lines in the ultraviolet region, and they all fitted his predictions within better than 0.1%. His formula seemed completely vindicated; he was equally correct in his prediction that all higher lines must crowd more and more together as  $n \rightarrow \infty$ , for according to his formula, the shortest wavelength ( $\lambda$  at  $n = \infty$ ) is 3645.6 Å, the so-called *convergence limit*.

Increasingly improved techniques in spectroscopy soon made it possible to record more and more of the ultraviolet region, as for example by photographing the emission spectrum of the hot gases in the sun's chromosphere during an eclipse; the 35 consecutive "Balmer" lines so recorded were all in good agreement with his formula. The convergence limit came to 3645.81 Å. Figure 22-14 represents a Balmer spectrum on a plate sensitive into the ultraviolet region.

**PROBLEM 22-6.** (a) Check Balmer's calculation for the fifth line,  $H_\epsilon$ , and for the convergence limit. (b) How many hydrogen lines should there be in the range between 3650.0 Å and 3660.0 Å?

But there was more. Balmer thought that on purely speculative grounds one might expect several complete new series of hydrogen lines for which  $\lambda = C \times \left( \frac{n^2}{n^2 - 3^2} \right)$ ,  $\lambda = C \times \left( \frac{n^2}{n^2 - 4^2} \right)$  and so on; and he also expressed the hope that this formula might guide the way to discoveries of series relationships among other spectra than that of hydrogen. It soon became evident that Balmer's suggestions were quite correct, although there was as yet no hint of a *physical* basis on which to arrive at this formula.

It is now customary to write Balmer's formula in a perfectly equivalent but more suggestive way which we shall adopt from now on:

$$\frac{1}{\lambda} = R \left( \frac{1}{2^2} - \frac{1}{n^2} \right), \quad (22-3)$$

where the value for the constant  $R$  is, according to modern measurements,  $(109,677.76 \pm 0.05) \text{ cm}^{-1}$ . The letter  $R$  has been chosen to honor the Swedish spectroscopist J. R. Rydberg (*Rydberg constant*  $R$ ).

PROBLEM 22-7. Derive Eq. (22-3) from (22-2), and show that  $R = 4/C$ .

Now we can restate Balmer's prediction of other series of hydrogen lines, by writing speculatively  $\frac{1}{\lambda} = R \left( \frac{1}{3^2} - \frac{1}{n^2} \right)$ ,  $\frac{1}{\lambda} = R \left( \frac{1}{4^2} - \frac{1}{n^2} \right)$ , etc., perhaps also  $\frac{1}{\lambda} = R \left( \frac{1}{1^2} - \frac{1}{n^2} \right)$ . In general, all these possible series can be summarized in one formula, namely,

$$\frac{1}{\lambda} = R \left[ \frac{1}{(n')^2} - \frac{1}{(n'')^2} \right], \quad (22-4)$$

where  $n'$  is a fixed integer (1,2,3 . . .) that distinguishes one series from the other (for the Balmer series,  $n' = 2$ ) and where  $n''$  also represents integers (namely,  $n'' = n' + 1, n' + 2, n' + 3 . . .$ ), but those integers that distinguish one line from the next in a given series, so that for  $H_\alpha$  and  $H_\beta$ ,  $n''$  is 3 and 4 respectively.

And indeed, in 1908, Paschen in Germany found two new hydrogen lines in the infrared whose wavelengths were given by the first of the speculative formulas or, in terms of Eq. (22-4), by  $n' = 3$  and  $n'' = 4$  and 5. Many other lines in this series have been identified since.

Table 22-3

Name of series	Date of discovery	Values in Eq. (22-4)
Lyman series	(1906-1914)	$n' = 1, n'' = 2, 3, 4 . . .$
Balmer series	(1885)	$n' = 2, n'' = 3, 4, 5 . . .$
Paschen series	(1908)	$n' = 3, n'' = 4, 5, 6 . . .$
Brackett series	(1922)	$n' = 4, n'' = 5, 6, 7 . . .$
Pfund series	(1924)	$n' = 5, n'' = 6, 7, 8 . . .$

To the Balmer series and the Paschen series there have gradually been added others, as by the improvement of experimental apparatus

and techniques new spectral regions could be explored. In the above list, the name of the series corresponds to its first observer, except, of course, that the *existence* of Balmer's lines were common knowledge at the time he formulated the series expression.

PROBLEM 22-8. Find the wavelength of the third line (always counting up from the longest wavelength) in the Lyman series, and the convergence limit for that series.

PROBLEM 22-9. Draw a long line and divide it into 20 equal intervals, about 1 cm apart. Let this represent a wavelength scale from 0 to  $80,000 \times 10^{-8}$  cm. On this scale enter the first, second, third, and last line of each of the five series. Do they overlap? Of how many series could one conceive between  $\lambda = 0$  and  $\lambda = \infty$ ? How many lines *in toto*?

The rest of Balmer's initial suggestions began to be fulfilled even sooner than those concerning other series for hydrogen emission; his formula did indeed prove suggestive in the search for relationships among the spectra of other gases. In 1896 E. C. Pickering observed a series of hitherto unknown lines in the absorption spectra of a star that virtually coincided with that of the Balmer series for hydrogen except that there was an additional line between every couple of Balmer-type lines. The whole set of wavelengths fitted the relation

$$\frac{1}{\lambda} = R \left[ \frac{1}{(n'/2)^2} - \frac{1}{(n''/2)^2} \right], \quad (22-5)$$

where  $n' = 4$  and  $n'' = 5, 6, 7 \dots$  (later, other series of this type were found, e.g., one with  $n' = 2$  and one with  $n' = 3$ ). The thought was voiced at first that this was evidence for some celestial modification of common hydrogen. The true solution to this intriguing puzzle was one of the earliest triumphs of the new model of the atom to which the historical development is here gradually leading us; it is such a good story that we will save it for the last chapter.

While Balmer's formula itself was not found to serve directly in the description of other spectra, it inspired formulas of similar mathematical form that were very useful in discerning some order in portions of a good many complex spectra. The Rydberg constant itself also reappeared in such empirical formulas. It became more and more clear that somehow the same physical mechanism had to be at work behind the variety of spectroscopic observations, although simultaneously the exciting search threatened to bog down under the sheer weight of data. But as we know now, it might almost have been equally good for the progress in this field if for a genera-

tion after Balmer no line spectrum analysis had been attempted, because the key pieces for the construction of the necessary conceptual scheme had to come first from an entirely different direction. And we now turn to follow that path.

### Further References

O. H. Blackwood, *et al.*, *An Outline of Atomic Physics*. New York: John Wiley (1937). Chapter V on spectroscopy and Chapter XV on the applications of spectroscopy in astrophysics. Highly recommended.

H. Buckley, *op. cit.* (Ch. 15). Chapters IV and VI.

H. Crew, *op. cit.* (Ch. 2), pp. 331-365.

P. Lenard, *op. cit.* (Ch. 9). Biographies of Fraunhofer, Kirchhoff, Bunsen, Maxwell, Hertz.

W. F. Magie, *op. cit.* (Ch. 2), pp. 354-365. Excerpts from Kirchhoff and Balmer; pp. 528-538, 549-561, excerpts from Maxwell and Hertz. Also W. S. Knickerbocker, *op. cit.*, Chapter XXXIII; H. Shapley and H. E. Howarth, *op. cit.*, pp. 196-201 (Fraunhofer), 276-282 (Kirchhoff); F. R. Moulton and J. J. Schifferes, *op. cit.*, pp. 450-465 (Maxwell, Hertz).

L. W. Taylor, *op. cit.* (Ch. 5). Chapters 36 and 37 on spectroscopy, including measurement of wavelengths with gratings.

W. P. D. Wightman, *op. cit.* (Ch. 6). Chapters XIII, XIV, XXII and XXIV.

## CHAPTER 23

### THE QUANTUM THEORY OF LIGHT

**23-1 Planck's empirical emission formula.** We must now pick up the story at the point where we left it in Section 22-4, where the classical conceptual scheme for the emission of light failed to account for the observed distribution of intensity *vs.* wavelengths in continuous emission spectra of solids and liquids. In October 1900, the German physicist Max Planck (1858-1947) (Fig. 23-1) disclosed that he had found for continuous emission spectra from ideal (blackbody) radiators an equation somewhat analogous to the one Balmer had given for the line emission spectrum of hydrogen; that is, Planck had discovered the one mathematical expression for representing the experimental energy distribution curves at every temperature (Fig. 22-7). One way of formulating his result is this: If we desig-



*Dr. Max Planck*

FIG. 23-1. Max Planck (1858-1947).

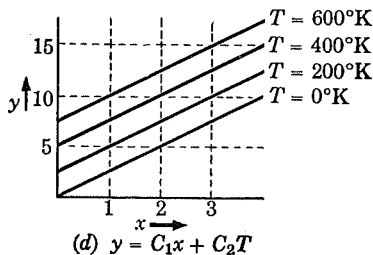
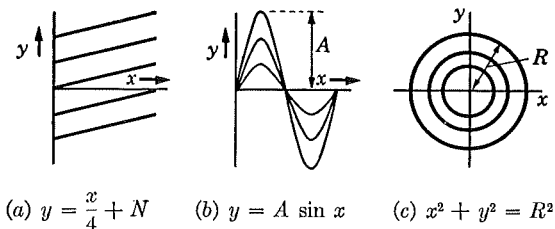


FIG. 23-2.

nate by the symbol  $E_\lambda$  the radiation energy emitted per second, per unit area of emitter surface, per unit wavelength interval, then

$$E_\lambda = \frac{C_1\lambda^{-5}}{e^{C_2/\lambda T} - 1}, \quad (23-1)$$

where  $C_1$  and  $C_2$  are constants, and  $e$  is a number (about 2.718).

But Planck, one of the greatest physicists of his time, was not at all satisfied with this accomplishment. After all, one can generally find a formula to represent any family of curves. For example, Fig. 23-2 shows three different families of curves—a set of straight lines, a set of sinusoidal waves, and a set of concentric circles, each drawn in the plane of a rectangular  $x$ - $y$  coordinate system, and each family identified also by the corresponding single analytic expression which would serve to plot the whole set.

**PROBLEM 23-1.** In Fig. 23-2(d) is shown a similar family of lines as in (a), but the coordinates have been given numerical values; we might have here a set of experimental data, simpler than but quite analogous to Fig. 22-7, possibly from some other branch of physics. Check by calculation for some points on each line that *every* line in this family is represented by the same equation  $y = C_1x + C_2T$ , where  $C_1$  and  $C_2$  are constant, with the same value for each line. Compute the numerical values of  $C_1$  and  $C_2$ .

Finding Eq. (23-1) was to Planck no more of an explanation of the physical mechanism responsible for the phenomenon than was Balmer's formula. Planck was, in fact, then working on the problem of deriving the energy distribution curve by assuming that light was emitted from a Maxwell-Hertz type of oscillator. In his autobiographical notes he relates that the problem appealed to him deeply because of the feature which we have had cause to marvel at—that the continuous spectrum from (ideal) emitters is entirely independent of their chemical properties. We are reminded of the law of universal gravitation, which also was found to hold regardless of the chemical nature of the participating masses. Here again there seems to be a universal law at work. In Planck's own words, this fact concerning emission "represents something absolute, and since I had always regarded the search for the absolute as the loftiest goal of all scientific activity, I eagerly set to work." We note parenthetically the philosophic kinship between Planck and Newton; the tone here expressed distinctly forms a counterpoint in contemporary philosophy of science.

As had all others who struggled with the emission problem, Planck failed at first, but at any rate his empirical formula of intensity *vs.* wavelength was a startling success. The very night of Planck's public announcement in a scientific meeting, a colleague extended the measurements beyond the range and accuracy previously obtained, and confirmed a complete agreement between the precise facts and Planck's formula—a case rather similar to Balmer's. Spurred on by this success, Planck once more set to work to provide a conceptual scheme from which his formula would follow as a necessary consequence—"until after a few weeks of the most strenuous labor of my life the darkness lifted and a new, unimagined prospect began to dawn." December 14, 1900, when Planck presented his solution, marked the effective birthday of contemporary physics.

**23-2 The quantum hypothesis.** Hertz's work had convinced Planck that radiation must be explained by the action of submicroscopic electric oscillators.\* The wall of a glowing solid emitter may thus be imagined as populated by such oscillators (perhaps the vibrating electrons, but this need not be specified), all at different frequencies. But how does such an oscillator gain energy? One mechanism might be the incessant collisions with neighbors as heat is supplied to the wall to keep it emitting. And how is light radi-

---

\* Planck referred to them as *resonators*.

ated? By sending out electromagnetic waves during the vibration, in varying amounts of energy and at the frequency of the particular oscillator. But just here Planck finds he must make a radical change in the classical picture: to fit observed facts, one must *not* assume, as seemed matter-of-course until then, that an oscillator of some frequency  $f$  can send out any part of its total amount of energy, no matter how small. On the contrary, the energy content of the oscillator, its kinetic and potential energy at any moment, can only be an integral multiple of the quantity  $hf$ , where  $f$  is the frequency of vibration and  $h$  is a universal constant, known as Planck's constant, whose value remains to be found by experiment. For example, the total energy at any one instant may be zero,  $hf$ ,  $2hf$ , . . .  $nhf$  ( $n =$  any integer), but never  $1.85hf$  or the like. That is to say, the energy of the oscillator is *quantized*, limited to integral multiples of one given amount or *quantum of energy*,  $hf$ . If ordinary pendulums, on losing their energy by friction during vibration, behaved according to such a rule, their amplitudes of vibration would not fall off gradually, but would decrease in a stepwise manner, each step representing an energy interval proportional in magnitude to the frequency of vibration.

A more concise rendition of the fundamental ideas of quantum theory can be given in terms of the concept *energy levels*: If the frequency  $f$  of an oscillator is constant despite changes in amplitude, then we might represent its energy content over a short time interval by a graph like that in Fig. 23-3. Each of these steps or changes of levels is occasioned by absorption or emission of energy correspond-

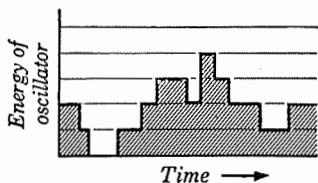


FIG. 23-3. The energy content which a single Planckian oscillator may have changes discontinuously from one level to the next.

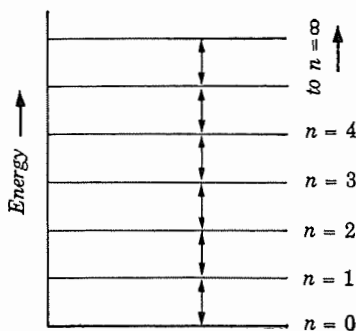


FIG. 23-4. Energy level scheme for one oscillator of frequency  $f$ . For any level, the energy is  $nhf$ ; transitions are permitted between neighboring levels, as indicated.



ing to a single quantum of magnitude  $hf$ . If we now draw all the different energy levels which the oscillator may reach at one time or another (Fig. 23-4), we obtain a ladder with equal steps. At any moment, the oscillator, as far as its energy content is concerned, is "located" on one or another of these energy levels, if it is not just then changing levels. Planck knew nothing about the physical nature of the oscillator, or about its actual behavior in time; therefore the energy level diagram is the best representation of oscillators, for it avoids premature or erroneous pictorializations of the physical situation within the emitter.

To this point we have been looking only at a single oscillator of frequency  $f$ . Now we extend our view to encompass all the uncountable oscillators of different frequencies on the surface of the glowing body. This body as a whole, in Planck's view, emitted the continuous spectrum of electromagnetic (light) waves by virtue of containing simple oscillators of all frequencies, each emitting light of a frequency  $\nu$  identical with its own frequency of vibration  $f$ . The energy  $E$  of each oscillator being quantized according to  $E = nhf$ , emission of energy can occur only during a sudden change in the amplitude of oscillation, for this amplitude is directly related to the energy content.

**PROBLEM 23-2.** A small object of mass  $m$  at the end of a spring oscillates about an equilibrium position with simple harmonic motion and an initial amplitude  $x_1$ . Recall (Section 17-6) that the potential energy of the system in that state of extension or compression is  $\frac{1}{2} \kappa x_1^2$ ; here  $\kappa$  is a constant indicating the stiffness of the spring, i.e.,  $\kappa =$  the "spring constant," the amount of force (dynes) required per cm extension. If in a particular case  $m = 3$  gm,  $\kappa = 10^2$  dynes/cm,  $x_1 = 4$  cm, find answers to the following: (a) What is the potential energy of the oscillator at the extreme ends of the vibration? When it passes through equilibrium position? What is the kinetic energy in these two cases? What is the total energy at all times? (b) In problem 5-21 you derived the period  $T$  of a spring, to be given by  $T = 2\pi\sqrt{m/\kappa}$ . What is the frequency  $f$  of vibration for this particular spring? (c) If the total energy of this *hypothetical* oscillator can only change in quanta of  $Hf$ , where by experiment  $H = 218$  erg-sec, compute its energy if it were to gain a quantum, or lose one quantum. What is the amplitude of vibration in each of these two cases? (d) Draw a graph for the first few energy levels corresponding to Fig. 23-4.

To Planck's fundamental postulate that the electric oscillator can only change its energy between neighboring energy levels  $E''$  and  $E'$  according to  $\Delta E = E'' - E' = hf$ , we must add the requirement

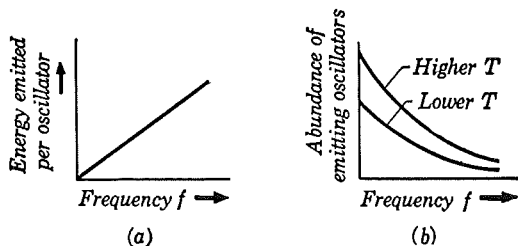


FIG. 23-5.

that at any moment the likelihood is much greater that the emitting body contains more elementary oscillators at low energies (low  $f$ ) than at high energies. From assumptions such as these, fully and quantitatively developed in a brilliant series of publications, Planck was able to deduce the emission formula [Eq. (23-1)] which he had previously found empirically, and he derived also the other well-established characteristics of continuous emission spectra, as for example Wien's displacement law.

While the details of the argument are on a most advanced level of mathematical physics, we may, in a qualitative way, make Planck's work plausible by means of the following curves. Figure 23-5(a) shows the energy that is emitted by individual oscillators during a drop in energy level, as a function of the frequency of vibration or of emitted light; it is a straight line graph, according to the assumption  $\Delta E = hf$ , and so shows that the high frequency radiation carries away the largest energy during emission, whereas low frequency emission does not deplete the emitter of much energy. On the other hand, the *number* of high frequency oscillators present in an emitter is postulated to be much smaller than the number of low frequency oscillators; see Fig. 23-5(b), where the relative abundance is shown in a schematic way for two temperatures; a rise in the temperature of the emitter, of course, brings into play more oscillators at all frequencies. Now if the assumptions in these two graphs be granted, we can at once make a prediction of the distribution curve for the *total* energy by multiplying the ordinates (vertical coordinates) of these two graphs at each frequency to obtain a new graph. The new ordinate will be (energy/oscillator)  $\times$  (number of oscillators at given frequency), i.e., total energy radiated at (or within a small frequency interval around) a given frequency; the abscissa (horizontal coordinate) will still be frequency. When such a graph is made, two things become apparent. The maximum intensity of radiation

must be in the range between the small and high frequencies, and for higher temperatures the peak of intensity shifts to the right, toward higher frequencies, i.e., shorter wavelength, exactly as expected by Wien's law. In short, the graph derived from Planck's quantum hypothesis exhibits the main features of the corresponding graph (Fig. 22-7) obtained by experiment.

PROBLEM 23-3. (a) Copy Fig. 23-5(a) and (b), superimpose on each a grid of vertical and horizontal lines for reference purposes, and construct the total energy distribution graph (for both temperatures). (b) Note that in Fig. 22-7, with which our new graph is to be compared, the abscissa is not  $f$  but  $\lambda$ . Therefore draw a second graph from the data for the first with the same quantity as ordinate but  $\lambda$  as abscissa (recall  $\lambda = c/f$ ). Compare its shape with Fig. 22-7.

From his fundamental assumptions, Planck deduced the following precise mathematical relation between the energy radiated ( $E_\lambda$ ), the wavelength or frequency of the radiation ( $\lambda$  or  $\nu$ ) at which the energy  $E_\lambda$  is measured, and the temperature ( $T$ ) of the whole (ideal, blackbody) emitter:

$$E_\lambda = \frac{8\pi\lambda^{-4} h\nu}{e^{h\nu/kT} - 1}, \quad (23-2)$$

where  $k$  = Boltzmann's constant (see Section 20-5), and  $h$  is the constant of proportionality that relates the smallest possible energy change of an electric oscillator, namely the quantum of energy, to its frequency of oscillation  $f$ . While we cannot follow Planck's mathematical derivation in detail, Eq. (23-2) will seem more plausible if we accept that, on the basis of the quantum hypothesis, (a) the average energy per degree of freedom of the emitter was not  $kT$ , as in the classical theory based on the principle of equipartition,\* but is given rather by the expression

$$\frac{h\nu}{e^{h\nu/kT} - 1};$$

and that (b) the number of degrees of freedom of the emitter at wavelength  $\lambda$  is  $8\pi\lambda^{-4}$ . Multiplying the last two expressions, we obtain Eq. (23-2).

By inspection, we can reach at once two conclusions:

(a) The theoretically derived Eq. (23-2) is indeed perfectly equivalent to the empirical Eq. (23-1), for if we replace  $\nu$  by  $c/\lambda$ ,

\* Each vibrating oscillator has, in general, potential and kinetic energy, and  $\frac{1}{2} kT$  is assigned for each of these two modes of carrying energy.

where  $c$  = velocity of light, we can identify  $C_1$  with  $8\pi ch$  and  $C_2$  with  $ch/k$ , the quantities  $c$ ,  $h$ , and  $k$  being universal constants.

(b) In the last statement it is implied that we may use the relations there to calculate  $h$ . Given the measurements of  $E_\lambda$ ,  $\lambda$ , and  $T$ , and knowing the factors  $c$  and  $k$ , there remains only  $h$  as unknown. On solving, we find

$$h = 6.62 \times 10^{-27} \text{ erg-sec.}$$

Planck's own calculations of  $h$  at the time gave him a value correct to within 1% of the present well-confirmed figure. It is one of the most fundamental constants in atomic physics, of the same importance as  $G$ , of the law of universal gravitation, is in astrophysics, or as  $N$  is in chemistry. The quantity  $h$ , initially evaluated as discussed above, made its appearance thereafter in many other connections, so that there are now several independent methods of calculating its value.

**PROBLEM 23-4.** An atomic oscillator sends out light at 7000 Å (red). At what frequency does the oscillator vibrate? What is the quantum of energy for this oscillator? What is it for another oscillator that emits violet light (at 4000 Å)? *Note well that the quantum of energy depends on the frequency of the oscillator; energy is not quantized, like electrical charge, in uniformly large bundles under all conditions.*

**PROBLEM 23-5.** An atomic oscillator sends out light at 5000 Å. Draw some portion of its energy level diagram (to a stated scale) according to the quantum hypothesis. What is the magnitude between successive steps on the diagram?

**PROBLEM 23-6.** In problem 23-2, the value of the quantity  $H$  was chosen quite arbitrarily. In actual fact, all oscillating systems, even springs, are considered to obey the rule that the smallest change in energy can proceed only according to the equation  $\Delta E = hf$ , where  $h$  = Planck's constant,  $f$  = frequency of vibration. Given the same spring as before, with the same mass  $m$ , stiffness  $\kappa$ , and initial amplitude  $x_1$ , calculate now the actual change of amplitude when the energy changes by one quantum. Why does the decay or increase of the oscillatory motion of springs, pendulums, and the like generally seem to us continuous when it is really quantized?

We might think that these successes must have gratified Planck deeply. However, he later confessed that his own postulate of energy quantization, to which he had been driven by the facts of emission, always remained to him deeply disturbing. He had indeed attained his goal of finding an "absolute" as nearly as man can—that is, he had found a universal and basic process—but at the cost of setting aside the equipartition theorem that until then had

been regarded as a generally applicable guiding concept. Furthermore, like most physicists of the time, he was still fundamentally convinced that natural processes are continuous; as Newton had expressed it, *natura non saltus facit*. Indeed, both the ancient faith in a sequence of cause and effect and even the usefulness of the mathematical calculus itself, seemed to depend on the proposition that natural phenomena do not proceed by jumps.

Not unexpectedly, the quantum hypothesis was regarded as rather too radical and unconvincing by almost all physicists at the time. Planck relates that he himself spent years trying to save science from discontinuous energy levels while the quantum idea "obstinately withstood all attempts at fitting, in a suitable form, into the framework of the classical theory." From a broader point of view, the general initial hesitation to plunge into quantum theory is only symptomatic of a useful law of inertia in the growth of science. One had to scan carefully the full implications of a step into such strange territory.

Moreover, Planck's initial treatment was not free from ambiguities that had to be resolved. There was, for example, a subtle paradox in Planck's reasoning. On the one hand, he jealously held on to the Maxwellian wave theory of light, considering it unquestionable that the light energy given off by a quantized oscillator spreads continuously over an ever-expanding set of wave fronts; and on the other hand, he implied that the subatomic oscillators, in contradiction to the classical theory, would radiate energy *not* while oscillating with a certain amplitude, but only while suddenly decreasing that amplitude. Lastly, Planck's work did not contain a physical picture or model for explaining such puzzles as the difference between continuous and line emission spectra.

If one chose, one could regard Planck's scheme as a formalism which at best left the "real" questions of physical mechanisms unanswered. The general acceptance of Planck's work depended therefore not just on experimental verifications, but on further demonstrations of the usefulness, even the necessity, of the quantum hypothesis in scientific thought. For some years the quantum hypothesis was uneasily suspended between doubt and neglect. Then Einstein extended and applied Planck's conception in rather the same manner that Avogadro developed Dalton's, and thereafter the doubts slowly began to fall away. Because Einstein's specific contribution to this topic depended on the clarification of a long-standing puzzle, the photoelectric effect, we shall first trace out the relevant facts.

**23-3 The photoelectric effect.** Ironically, the first facts leading to the photoelectric effect, and through that eventually to the recognition that the classical theory of light had to be fundamentally revised, were by-products of the very same research that seemed at the time to furnish the most convincing proof for the classical theory—namely, Hertz's experiments on the propagation of electric waves. As we noted, Hertz mentioned that an electric spark would jump more readily between the metal spheres if they were freshly polished. It soon became clear, through the work of various contributors, that the ultraviolet component of light—as from a spark itself—had the effect of drawing or expelling negative charges from the clean surface of metals, and that it was these charges that aided in maintaining a suitable transient current between the spheres in Hertz's equipment. Indeed, some metals, notably the alkali metals, exhibited the same effect even with visible light. The air around the metal was not involved; on the contrary, when two clean metal plates (Fig. 23-6) are sealed in a transparent, well-evacuated quartz tube and are connected to a battery which charges one of the plates negatively, the latter, if illuminated, copiously emits negative charges that are then attracted to the positive plate and recirculated by way of the battery. The current so obtained clearly deserves the name *photoelectric current*.

The exact nature of those negative charges was not clarified until 1897. At that time, principally through the work of J. J. Thomson, electricity was found to be carried in fixed quantities of charge by discrete particles much smaller than the lightest atoms. This identification of electrons was first made from a study of the current through an almost completely evacuated tube when a sufficiently high potential difference was applied between two electrodes. Although the electron concept was eventually found to be essential in explaining every phenomenon involving electric charges, the free electron was initially a radical hypothesis, in serious contradiction to the classical view of the structure of matter, for it had previously not been necessary to entertain the notion that any material particle smaller than the

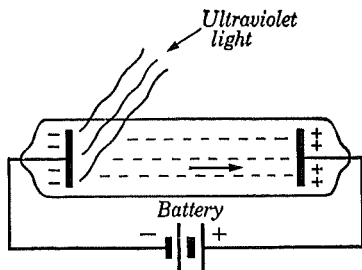


FIG. 23-6. A photoelectric current is set up through the circuit when ultraviolet light falls upon the electrode at the left.

atom should exist by itself. J. J. Thomson himself wrote:

“At first there were very few who believed in the existence of these bodies smaller than atoms [i.e., the electrons]. I was even told long afterwards by a distinguished physicist who had been present at my lecture at the Royal Institute that he thought I had been ‘pulling their legs.’ I was not surprised at this, as I had myself come to this explanation of my experiments with great reluctance, and it was only after I was convinced that the experiments left no escape from it that I published my belief in the existence of bodies smaller than atoms.”

However, the electron quickly explained and coordinated a wide variety of phenomena. The flow of electrons accounted not only for currents in cathode-ray tubes and photoelectric cells, but also in solid metallic conductors; in a completely evacuated tube, a heated metal wire (filament) “boils off” electrons which are attracted by a positively charged plate nearby and can be recirculated by a battery connected between plate and filament (*thermionic* tube, or diode as used in radio circuits); the overabundance or deficiency of one or more electrons would turn an atom into a negatively or positively charged ion, and so help to account for the facts of electrolysis; the electrons associated with a neutral atom, one in which the positive and negative charges balance, could be identified more definitely with the oscillators responsible for the emission of light on the classical (and later on the quantum) model of radiation.

More detailed studies of the photoelectric effect itself revealed the following experimental facts:

(a) When, in the apparatus shown in Fig. 23-6, the light incident on the left-hand plate is of one given frequency (“monochromatic” light), the number of electrons (“photoelectrons”) emitted per second per unit area of plate, as measured by the current through the tube and the wires, increases with the intensity of the incident light. This is not unexpected, for the energy of the incident light must somehow be absorbed by the electrons as they leave the cathode; in this conversion from light energy to the energy in an electric current, an increase in the former should be attended by a corresponding increase in the latter.

(b) When photoelectrons of mass  $m_e$  are “knocked out” from the metal by the incident light, they must leave with some initial kinetic energy  $\frac{1}{2} m_e v^2$ . To find this energy experimentally, we simply remove the battery connections in Fig. 23-6 which put an accelerating poten-

tial between the electrodes, and apply instead a potential difference in the opposite sense, a *retarding voltage*  $V_r$  (see Fig. 23-7). Under these conditions only those electrons will reach the right side whose initial kinetic energy  $\frac{1}{2} m_e v^2$  is as large as or larger than  $V_r e$ . In such an experiment we may increase  $V_r$  until no photoelectric current flows at all. At that point the "retarding voltage" has become a "stopping" potential,  $V_s$ , that is, even those electrons emitted by the action of the light with the highest speed  $v_{\max}$  just fail to reach the plate; in equation form,

$$V_s e = \frac{1}{2} m_e v_{\max}^2. \quad (23-3)$$

We should expect to find that the maximum kinetic energy of photoelectrons increases in some way with the intensity of the incident light, just as high ocean waves will not only move more of the pebbles on the shore, but also give them greater speeds. However, the experiments showed that *the maximum kinetic energy of photoelectrons is not a function of the intensity of incident light of a given frequency*. This was unexpected. It implied that an electron in the cathode can draw from the incident light energy up to a fixed maximum only, rejecting the remaining energy if the light waves are very intense, and, if the light is very feeble, presumably accumulating that maximum energy before leaving the surface.

**PROBLEM 23-7.** In 1902 P. Lenard determined that photoelectrons released from a metal on which the ultraviolet portion of sunlight was falling needed a stopping potential of 4.3 volts. What was the maximum speed of these electrons? (Be careful to use consistent units.)

(c) It is natural to inquire next whether the  $KE_{\max}$  of photoelectrons depends on the frequency  $\nu$  of the incident light, and experiments showed, though at first only qualitatively, that the KE does in fact increase with frequency. But it was not a case of simple proportionality: for each metal there was some frequency of light  $\nu_0$  below which no photoelectric emission could be obtained no matter how strong the incident radiation. This *threshold frequency*,  $\nu_0$ , for zinc is about  $8.5 \times 10^{14} \text{ sec}^{-1}$  ( $\lambda_0$  is about 3500 Å), and for metallic sodium is around  $5 \times 10^{14} \text{ sec}^{-1}$  ( $\lambda_0$  below 6000 Å). This observa-

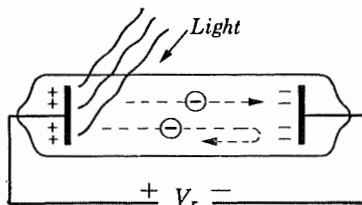


FIG. 23-7. To overcome the retarding potential  $V_r$ , the photoelectron must have a sufficiently large initial kinetic energy.



tion raised the same type of conceptual difficulties as noted in (b) above concerning the mechanism by which the electron absorbs energy from the light.

(d) Finally, it was also observed that whereas no amount of illumination below the threshold frequency would yield any photoelectrons from a given metal, even the feeblest intensity of light of a higher frequency than  $\nu_0$  sufficed to obtain photoelectric currents *without any delay*. Modern experiments have confirmed that the time interval between the arrival of light and the departure of electrons from the surface is less than  $3 \times 10^{-9}$  sec. That, however, is in flat contradiction to our expectations: we have measured already the energy photoelectrons may have, and we can calculate how long we should have to let light fall on a metal surface in a typical experiment before that much energy is available to the electron. One result of such a calculation is that an electron would have to wait for some seconds before it could accumulate the necessary energy from a weak beam of light—always assuming, as we do in the Maxwellian picture of light propagation, that the energy of a light wave is uniformly distributed over the whole wave front.

**23-4 Einstein's photon theory.** The resolution of these problems came in another 1905 paper of Einstein,\* entitled modestly, "On a heuristic point of view concerning the generation and transformation of light."\*\* Einstein (Fig. 23-8) begins by paying tribute to the wide usefulness of Maxwell's theory of light; it "has proved itself excellently suited for the description of purely optical phenomena [reflection, refraction, interference, polarization, etc.], and will probably never be replaced by another theory." But this classical theory of light was designed and tested above all with reference to the *propagation* of light, its progress through space measured over relatively long time intervals. Therefore, he continues, one

\* For a splendid biography of Einstein, see P. Frank, *Einstein, His Life and Times*, New York: A. Knopf (1947).

\*\* *Heuristic* = useful as an aid in the advancement of a conceptual scheme, but not necessarily in the final correct form; an example of a heuristic device is the use of analogies. It is curious to note that Newton, in communicating to the Royal Society in 1675 a corpuscular view of light, entitled his paper "An Hypothesis Explaining the Properties of Light," and wrote that this hypothesis was also a heuristic one: ". . . though I shall not assume either this or any other hypothesis, not thinking it necessary to concern myself whether the properties of light discovered by men be explained by this or any other hypothesis capable of explaining them; yet while I am describing this, I shall sometimes, to avoid circumlocution . . . speak of it as if I assumed it."



*A. Einstein.*

FIG. 23-8. Albert Einstein in 1912.

should not be surprised if contradictions with experience arise when the investigation turns to problems where everything is contingent on a *momentary* interaction between light and matter. Two prominent examples of the latter case where troubles have developed are the origin or emission of light from an oscillator and the transformation of light energy into the kinetic energy of photoelectrons.

Now Einstein proposes the following conceptual scheme: *The energy of light is not distributed evenly over the whole wave front, as the classical picture assumed, but rather is concentrated or localized in discrete small regions* or, if we may say so, in “lumps” or “bundles.” A mental image is difficult to come by, but this analogy might help: If the same statement could be applied to the energy of water waves, we would expect that an incoming wave could move a floating cork up and down (i.e., give the cork energy) only at certain spots along

the wave front; between those spots, the passing wave would leave the cork undisturbed.

One can see the roots of Einstein's idea in Planck's work. We remember that the Planckian oscillator can lose energy only in quanta of  $hf$ , and that the light thereby emitted has the frequency  $\nu = f$ . But in introducing a new concept of light emission Planck did not suggest in the least that a change is also necessary in the classical view concerning the nature and propagation of light waves after the moment of emission; the Maxwellian picture of expanding electromagnetic waves with evenly and continuously distributed energy remained effective. However, this raised conceptual problems of the following kind: if a Planckian oscillator can change its energy only in steps of  $hf$ , how can it absorb energy from light in which the energy is thinly distributed? If not by a slow continuous increase in oscillation, which is prohibited by the quantum theory, how can the oscillator store up the stream of energy given to it by incident light until there is enough for a jump to a higher energy level? Even if a storage mechanism could be imagined, the fact of instantaneous emission of photoelectrons [item (d) in Section 23-3] shows that no such storage actually occurs.

Roughly speaking, Einstein therefore extended the quantum concept to the energy in the emitted light. The quantum of energy ( $hf$  ergs) lost by the emitter does not spread out over the expanding wave front, but *remains intact as a quantum of light with energy  $h\nu$  ergs, where  $\nu$  of light =  $f$  of oscillator*. (To the quantum of light the name *photon* was later given.) Each wave front of light emitted by a glowing solid, or the sun, etc., must therefore be imagined to be studded with photons, like little dots on a balloon, traveling with the wave at the speed of light; and when the wave expands as it progresses, the light becomes weaker because the distance between neighboring photons increases. Ordinarily the light beams we observe in experiments on reflection, refraction, polarization, and the like make use of beams with substantial cross section; and then we are dealing with enormous numbers of photons on each wave front. Therefore the individuality of those photons is masked, just as an ordinary stone appears to our eyes to be made of continuous matter rather than of individual atoms. But in the absorption and transformation of light energy by an individual submicroscopic oscillator, as in the photoelectric effect, the finer "structure" of the light wave becomes important. That is why the classical theory of light remains completely valid and most appropriate for dealing with the ordinary problems of optics, whereas the quantum theory

of light must be used to understand the interaction between light and individual atoms.

An application of these fundamental ideas by Einstein yielded at once a quantitative explanation of the photoelectric effect. A quantum of light, a photon, falling on an atom may or may not give it enough energy—all in one “bundle”—to loosen an electron,\* if the photon's energy is enough, the electron can free itself by doing the work  $\mathfrak{W}$  that must be done against the force of electric attraction binding it to the atoms of the cathode; if the photon's energy is less than  $\mathfrak{W}$ , the electron will not be emitted. This at once explains why emission is instantaneous if it takes place at all [item (d)]. But now the meaning of the threshold frequency  $\nu_0$  is clear also: to break the electron from an atom at the surface, the photon energy  $h\nu$  must be at least equal to  $\mathfrak{W}$ , or

$$h\nu_0 = \mathfrak{W}. \quad (23-4)$$

Furthermore, we expect different atoms to have different internal structure, so that  $\mathfrak{W}$ , and consequently also  $\nu_0$ , should vary from one element to another, as indeed it does.

For light of higher frequency than  $\nu_0$  each photon has an energy  $h\nu$  larger than  $\mathfrak{W}$ , so that energy is left over to let the photoelectron leave with some initial speed. Because light can penetrate the top-most layers of atoms, the photoelectric effect may take place below the surface. If the photoelectron then has to pass through layers of atoms before reaching the surface of the metal plate, it may lose some of its speed; but some electrons will have suffered no such losses and will come off with the full speed  $v_{\max}$ , directly toward the anode. For these we write

$$h\nu = \frac{1}{2} m_e v_{\max}^2 + \mathfrak{W}. \quad (23-5)$$

This is Einstein's equation for the photoelectric effect, an application, on the level of individual atoms, of the law of conservation of energy to the quantum theory of photoelectricity. It contains the algebraic expression of the experimental fact noted in Section 23-3(b) and (c) that the  $KE_{\max}$  of photoelectrons increases with the frequency  $\nu$ , but not with the intensity of the incident light. On the other hand, a larger intensity means more photons and larger numbers of photoelectrons per second, in accordance with observation (a). Thus, Eq. (23-5) summarizes all of the observations in one mathematical law.

But again, it did more. We must realize that at the time of

\* Note that no picture is supplied for imagining how the photon energy is converted to the energy of the electron.

Einstein's proposal, the data on hand were incomplete and often only qualitative. Precise and reproducible measurements of the photoelectric effect are still most difficult; for example, roughness or small impurities on the metal surfaces will greatly affect the results for  $\nu_0$  and  $v_{\max}$ . Therefore Einstein's equation was essentially still a prediction rather than a summary of detailed quantitative facts. It was the fulfillment of these predictions in later experiments which gave Einstein's work such importance and changed the photon from a "heuristic point of view" to a fundamental part of contemporary scientific thought.\*

In Eq. (23-5), the constant  $\mathfrak{W}$  is often called the work function of the metal and  $h$  is, of course, Planck's constant. The equation indicates how both quantities can be found by experiment, for if we were able, by careful experimentation with two different frequencies of light  $\nu_1$  and  $\nu_2$ , to ascertain in terms of the necessary stopping potentials  $V_{s1}$  and  $V_{s2}$  what the factor  $\frac{1}{2} m_e v_{\max}$  is for the photoelectrons in each case [see Eq. (23-3)] then we can write

$$h\nu_1 = V_{s1}e + \mathfrak{W},$$

$$h\nu_2 = V_{s2}e + \mathfrak{W}.$$

Therefore 
$$h = (V_{s1} - V_{s2})e/(\nu_1 - \nu_2). \quad (23-6)$$

The constant  $\mathfrak{W}$  can be quickly calculated from Eq. (23-2) or (23-5) after  $h$  is obtained. And of course here is a chance to check the value for  $h$  from Eq. (23-2), i.e., from Planck's theory of energy quantization in the radiating oscillator, a theory supplemented but by no means invalidated by Einstein's work. The credit for making the necessary precise measurement under extremely difficult experimental conditions belongs to the same man who had previously measured the electronic charge  $e$ , namely, the American physicist Robert A. Millikan.\*\* The result, obtained in 1916, was a coincidence of the values of  $h$  by these two very different methods, within the experimental error.

**PROBLEM 23-8.** In a typical experiment, 10 ergs of light energy from a mercury arc are allowed to fall on one  $\text{cm}^2$  of surface of iron during each second. Clean iron reflects perhaps 98% of the light, but even of the remaining energy only about 3% falls into the spectral region above the thresh-

---

\* Nominally, it was mainly for this work on the quantum theory of light rather than for his relativity theory that Einstein was awarded a Nobel prize in 1921.

\*\* Recipient of the 1923 Nobel prize in physics.

old frequency of the metal. (a) How much energy is available for the photoelectric effect? (b) If all the energy in the effective spectral region were of wavelength 2500 Å, how many photoelectrons should be emitted per second? (c) What is the current in the photoelectric tube in amperes? (d) If the threshold frequency is  $1.1 \times 10^{15} \text{ sec}^{-1}$ , what is the work function  $\mathcal{W}$  for this metal? (e) What is the stopping potential (in volts) for photoelectrons emitted by light at 2500 Å?

PROBLEM 23-9. Sketch the graph you expect to obtain when plotting  $\nu$ , the frequency of light incident on a metal surface (abscissa), *vs.*  $\frac{1}{2} m_e v_{\text{max}}^2$  or  $V_s$  for the photoelectrons (coordinate). Should the line be straight? What is the physical meaning of the intercept on the horizontal axis? What is the equation for the line? What is the meaning of the slope of the line?

**23-5 The photon-wave dilemma.** Einstein's photon model of light was soon found fruitful in explaining other phenomena, e.g., photochemistry and anomalous specific heats, two topics to which Einstein himself greatly contributed. As has been aptly said:

"It is striking how thoroughly the theory affects our judgment of the importance of one field or another. Strictly from the experimental point of view, the photoelectric effect seems to be a rather remote corner of physics with technical applications of no great significance. However, Einstein's theory makes it clear that the photoelectric effect gives evidence of the nature of light, one of the greatest problems of physics." \*

By showing that the energy of light was quantized, Einstein contributed directly toward the acceptance of Planck's original quantum theory, then still without general support. But, a little like Dalton, who repudiated Gay-Lussac and Avogadro, Planck himself was far from pleased with Einstein's photon. By its acceptance, he wrote in 1910, "the theory of light would be thrown back by centuries" to the time when the followers of Newton and Huygens fought one another on the issue of corpuscular *vs.* wave theory. All the fruits of Maxwell's great work would be lost by accepting a quantization of energy in the wave front—and all that "for the sake of a few still rather dubious speculations." \*\*

\* From *Introduction to Atomic Physics*, by O. Oldenberg (1949).

\*\* It is amusing to note that the work of Thomas Young in 1801, which introduced Huygens' wave theory and became the basis for Maxwell's view, was generally attacked at the time by the proponents of the current corpuscular theory on the basis that the wave theory, as one critic put it, "can have no other effect than to check the progress of science and renew all those wild phantoms of the imagination which . . . Newton put to flight from her temple."

Indeed, the inconvenience of having to deal with photons is very real. They represent bundles of energy without having ordinary *rest mass*; in this they differ from the Newtonian *corpuscles* of light, leaving only a faint analogy between them, even though it is customary to refer to Einstein's photon theory as the *corpuscular* or *particulate* theory of light. But our mind tends to insist on a good picture, and it requires great self-control to visualize a quantum of energy without bringing in some matter to which to attach it. It was a little easier to think of Maxwell's light energy spread evenly through the "field" along a wave front. Then, too, there are other questions: How large in volume and cross-sectional area is the "spot" on the wave front where the photon is located? What could be the meaning of "wavelength" and "frequency" of light, which determine the energy content of the photon by  $E = h\nu$ , if the photon is, so to speak, only a dot on the wave front, and not part of a whole wave train? By what mechanism does the wave determine the path of the photon in such wave phenomena as interference and polarization? How does an electron absorb a photon?

In the past, one partial answer to such questions has been to hold concurrently two separate views on light—the wave and the photon model—and to apply one or the other as required by the problem. The more recent answer, which allows, in principle, a solution of every problem with physical meaning in this field, involves combining both views, and assuming the photons distributed over the wave front in a statistical way, i.e., not individually localized at a particular point. But at our level the more practical solution is, first of all, to realize that some of these questions, while possibly very disturbing, are asked on the basis of a mechanical view of atomic phenomena that may stem from an erroneous transfer of experience with large bodies obeying simple Newtonian laws. Thus the "size" of the photon is not a concept that we should expect to have the same meaning as the size of marbles and projectiles. Furthermore, photons (and atomic particles as well) differ from water waves, pebbles, and other large-scale entities in that one cannot make various experiments on the *same* subatomic entity. One can localize and measure and weigh a stone, find its velocity, etc., and all the while it is the same unchanged stone. But a photon, after it has been sent into a Geiger counter or a photographic emulsion, is no more; two photons on which we impress different experimental conditions—for example, in searching for wave and for corpuscular properties—are not, strictly speaking, the same entity.

As the physicist Max Born has said:

“The ultimate origin of the difficulty lies in the fact (or philosophical principle) that we are compelled to use the words of common language when we wish to describe a phenomenon, not by logical or mathematical analysis, but by a picture appealing to the imagination. Common language has grown by everyday experience and can never surpass these limits. Classical physics has restricted itself to the use of concepts of this kind; by analyzing visible motions it has developed two ways of representing them by elementary processes: moving particles and waves. There is no other way of giving a pictorial description of motions—we have to apply it even in the region of atomic processes, where classical physics breaks down.

“Every process can be interpreted either in terms of corpuscles or in terms of waves, but on the other hand it is beyond our power to produce proof that it is actually corpuscles or waves with which we are dealing, for we cannot simultaneously determine all the other properties which are distinctive of a corpuscle or of a wave, as the case may be. We can therefore say that the wave and corpuscular descriptions are only to be regarded as complementary ways of viewing one and the same objective process, a process which only in definite limiting cases admits of complete pictorial interpretation . . .” \*

However, no matter how uncomfortable the wave-photon duality, or how great the disparity between concept and intuition at this stage, the photon theory stands on its proven power to explain, to predict, to stimulate further discoveries. We may finally illustrate this power by brief references to a few examples.

**23-6 Applications of the photon concept.** (a) *Fluorescence.* A material is fluorescent if it absorbs light and immediately re-radiates it. It had long been noticed that in general the frequency of such re-radiated light is equal to or less than the original incident light. This is at once plausible when we imagine the following processes. The incident photon of energy  $h\nu$  raises the atomic oscillator to a new level of higher energy than before. There is a general tendency of all oscillators to remain only briefly at high energy levels and to spend their time mostly at the lowest levels; thus the oscillator

---

\* From *Atomic Physics*, by Max Born, New York: Hafner Publishing Co., 4th ed. (1946), p. 92.



quickly re-radiates the energy, though it may do so in two or more stages. If the absorbed photon is again given up in one process, we have resonance radiation (see Section 22-6), but if the original energy  $h\nu$  is to be divided, the frequency  $\nu_r$  of the re-radiated light must be lower.

PROBLEM 23-10. What does the above account reveal concerning the relationship between the frequency  $f$  of the oscillator and  $\nu$  of incident, absorbed light?

PROBLEM 23-11. In some cases light is re-radiated at a *higher* frequency. How can the model account for this?

(b) *X-ray photons.* In 1895, W. K. Röntgen discovered that a beam of cathode rays (soon after identified as high-speed electrons) could create a new type of radiation when allowed to fall upon an obstacle such as the glass of the tube itself. In time, these so-called x-rays were shown to be high-frequency electromagnetic radiation of the same fundamental nature as radio waves, gamma-rays, and all other forms of "light" (see Table 23-1).

Table 23-1. Spectrum of Electromagnetic Radiation

<i>Name of radiation</i>	<i>Approximate range of wavelengths</i>
Radio waves	a few meters and up
Microwaves	a few mm to meters
Infrared waves	7500 Å to 0.01 cm
Visible light	4000 to 7500 Å
Ultraviolet light	100 to 4000 Å
X-rays	0.1 Å to 500 Å
Gamma-rays	less than 0.5 Å

Nowadays, x-rays are generated by the bombardment of a metal target forming the anode in a hot-filament cathode-ray tube. In a sense this is the reverse of the photoelectric effect; now photons are "knocked out" from the substance by incident electrons. We may provisionally imagine that the energy of the incident electron is first absorbed by the atomic oscillators, and then is re-emitted as a photon when the oscillator returns to its original energy level. Therefore we expect that the energy of the x-ray photon is equal to that of the electrons in the beam, or

$$h\nu = \frac{1}{2} m_e v^2. \quad (23-7)$$

In practice, we find that Eq. (23-7) gives the highest frequency of x-rays obtainable for a given electron beam. Furthermore, when such an x-ray in turn is allowed to fall on a clean metal surface, it causes the emission of photoelectrons of about the same velocity as that of the original beam (the work function being very small compared with the energies involved in x-rays).

**PROBLEM 23-12.** Figure 23-9 shows schematically an apparatus for producing x-rays. If  $V_{ap} = 50,000$  volts, (a) what is the speed of the electrons incident on the anode, and (b) what is the energy of the x-ray photons emitted? (*Note:* this figure gives only the maximum energy, for the electron may not give all its energy to an oscillator.)

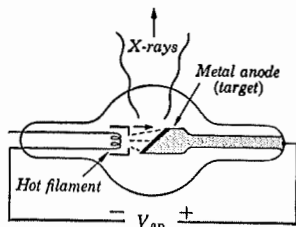


FIG. 23-9. X-rays are generated when a stream of high-speed electrons falls upon a target.

(c) *Particle-photon interaction.* As mentioned in Section 16-6, when a fast-moving electron passes by an atom, it may be deflected, and light (we would now say a *photon*) is emitted; the law of conservation of momentum is obeyed in this process if we can assign a momentum to the photon. But now we can understand the phenomenon that was previously only described as an empirical fact, for it follows from our discussion of the mass-equivalence of energy (Section 17-7) that one may assign a mass\*  $m = h\nu/c^2$  to a photon of frequency  $\nu$ ; and since momentum = mass  $\times$  velocity, the momentum associated with a photon is

$$\mathfrak{M}_{ph} = (h\nu/c^2) \times c = h\nu/c = h/\lambda. \quad (23-8)$$

This is exactly the amount of momentum needed to explain the experimental facts referred to in Section 16-6; that is, if we regard the momentum of the photon  $\mathfrak{M}_{ph}$  as a vector quantity of magnitude  $h/\lambda$  in the direction of the photon's motion, and if the velocity of the electron is  $\mathbf{v}_1$  before and  $\mathbf{v}_2$  after its change of motion near an atom, then we can write and experimentally verify the following equation:

$$m_e \mathbf{v}_1 = m_e \mathbf{v}_2 + \mathfrak{M}_{ph}. \quad (23-9)$$

\* Note that this is relativistic mass, not rest mass. If the photon had any rest mass  $m_0$ , what should be its observed mass at the usual speed of photons?

The momentum change of the atom in whose vicinity this process occurs is relatively small and has been neglected. The law of conservation of energy is obeyed at the same time, and can be written

$$\frac{1}{2} m_e v_1^2 = \frac{1}{2} m_e v_2^2 + h\nu. \quad (23-10)$$

In Section 16-6, we described Compton's experiment, which dealt with the reverse of the case just mentioned, namely, with the collision of an x-ray photon with an electron. In that case we may write the usual momentum and energy equation, and experiments again would completely bear out the implications:

$$(\mathcal{M}_{\text{ph}})_1 = (\mathcal{M}_{\text{ph}})_2 + m_e \mathbf{v}, \quad (23-11)$$

$$h\nu_1 = h\nu_2 + \frac{1}{2} m_e v^2. \quad (23-12)$$

**PROBLEM 23-13.** What is the mass, momentum, and energy of an x-ray photon of wavelength 1 Å?

**PROBLEM 23-14.** If we used x-rays of wavelength 1 Å and watched those that are scattered from electrons in a direction at right angles to the incident path, we would find that their wavelength after scattering is 1.024 Å, while the electron recoils as shown in Fig. 23-10. (a) Explain why there is a change in wavelength. (b) Calculate the speed and direction at which the electron recoils. (It is assumed throughout that the electron was initially stationary, and was not bound to an atom with any significant force.)

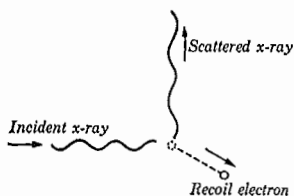


FIG. 23-10.

**23-7 Quantization in science.** The enormous change in the natural sciences during the one hundred years following 1808 can be correlated with a strangely similar discovery in each of such widely different fields as biology, chemistry, and physics—the discovery that there exists a structure, an elementary discrete quantization as it were, in the material on which each field builds. The case just discussed, that of energy quantization in systems that exhibit vibratory motion or periodic fluctuation (as in a Planck oscillator or in the electromagnetic field), is but the last of a series of such developments.

We recall that Dalton showed in the 1800's that chemical matter is not to be regarded as an infinitely divisible continuum, but that

it consists of atoms which maintain their integrity in all chemical transactions. In biology, the cell theory of Schleiden (1838) and Schwann revealed that all living matter consists of individual cells and their products. Gregor Mendel's work (1865) led to the immensely fruitful idea that the material governing heredity contains a structure of definite particles, or genes, which may be handed on from one generation to the next without change. Meanwhile the physics of heat, electricity, magnetism, and light, that is, that part of physics which the 18th century had visualized largely in terms of the action of imponderable fluids, was being rephrased in a similar way. Sensible heat was identified with the motion of discrete atoms. Electrical charge was found to be carried by the electron (whose motion would serve to set up magnetic fields, thereby reducing two topics to one). Finally, the energy of light and of atomic oscillators was found to be quantized also. In short, it was as if these new views in the various branches of science stemmed from a similarly directed change of the mental model used to comprehend fundamental quantities and processes, a change in the guiding idea from a continuum to a well-defined particle or quantum. In part, one can explain such changes by noting that the development of better instrumentation and techniques opened up a new level of observation. But this is largely arguing *a posteriori*, for such technical developments frequently were stimulated by the proposal of the idea which they later helped to fortify. One must allow the possibility of a change of tone and mood in the whole field, of a general conceptual movement not related to any single cause—yet another expression of the organic and dynamic nature of science.

#### Further References

O. H. Blackwood *et al.*, *op. cit.* (Ch. 22), Chapter IV on blackbody radiation and quantum theory.

H. Buckley, *op. cit.* (Ch. 15). Chapter X.

A. Einstein and L. Infeld, *The Evolution of Physics*. New York: Simon and Schuster (1938). Particularly part IV. Excellent reading.

O. Oldenberg, *Introduction to Atomic Physics*. New York: McGraw-Hill (1949). Part IV on quantum theory of light.

L. W. Taylor, *op. cit.* (Ch. 5), pp. 808-814.

## RADIOACTIVITY AND THE NUCLEAR ATOM

**24-1 Isotopes.** The account to this point has given us an atom that has some well-defined and some still mysterious features. Around 1910, just about one century after Dalton's announcement of the atomic theory of chemistry, one could be fairly certain of a few facts. For example, the size of atoms was known, though only very roughly, from kinetic theory; and while the internal structure of the atom was in doubt, it was certain that some of the components had to be charged, for electric currents of all kinds needed the electron, and all theories of light agreed on electric oscillators in the atom. Also, it had been shown by Lenard in 1895 that a stream of cathode rays (electrons) would easily penetrate foils or sheets of material and, although very thin, such foils still represented many thousand layers of atoms, so that the fact of penetration led one to think of the atom no longer as a ball-like solid, but as a porous, open structure.

A new set of astonishing facts was coming on the scene subsequent to the discovery of radioactivity in 1896 by Henri Becquerel. Atoms of the heavier elements such as uranium, thorium, and radium were found to disintegrate spontaneously by the emission of either a high-speed  $\alpha$ -particle (with a positive charge numerically equal to that of two electrons and a structure that was later found to be that of the helium nucleus), or a high-speed  $\beta$ -ray (a particle with the same rest mass and negative charge as any ordinary electron). In many elements, the emission of either particle was attended by that of a  $\gamma$ -ray, a very penetrating type of short-wavelength electromagnetic radiation.

The most central facts concerning natural radioactivity, which will serve as background for the following discussion, are not difficult to summarize in modern terms. Above all, an element (e.g., uranium) as prepared by a chemist is not the simplest starting point of investigation for radioactivity, because such a sample consists generally of a mixture of isotopes (atoms of the same chemical behavior but differing from one another in atomic weight\*) in natural

---

\* This is contrary to the most fundamental assumptions of Dalton and all chemical thinking in the 19th century. Does this discovery invalidate the work of the chemists?

radioactivity, and in the reaction to nuclear bombardment. This distinction has previously been referred to in Section 20-5 in connection with the separation of uranium isotopes by diffusion of uranium hexafluoride through porous walls. Another explicit application of the concept was made in Chapters 16 and 17, where we traced several nuclear reactions.

All our references to isotopes so far have emphasized their differences in weight, and this is the difference on which was based the first separation of isotopes by J. J. Thomson in 1912. If the gas or vapor from one element is ionized so that the individual atoms carry a charge, a beam of such ions can be deflected by an electric field, a magnetic field, or both applied simultaneously. But the amount of deflection depends on the mass of the ions. A brief hint of the operation of an isotope separator (*mass spectrograph*) working on such a principle will suffice.\* Figure 24-1 represents an ion beam emerging with initial speed  $v$  from the box at the left into the gap between plates  $A$  and  $B$ , held at a potential difference  $V_{AB}$ . If each ion has the same (negative) charge  $q$ , each will experience a constant force  $F_{\text{net}} = (V_{AB}q)/d$  and an acceleration  $a = F_{\text{net}}/m$  while traversing the distance  $l$ . Just as in the case of a projectile in a constant gravitational field and with horizontal initial velocity, the displacement  $s_y$  of the ion during the time of transit  $t = l/v$  under the action of force  $F_{\text{net}}$  is

$$s_y = \frac{1}{2} at^2 = \frac{1}{2} \frac{F_{\text{net}}}{m} t^2 = \frac{1}{2} \frac{V_{AB}q}{dm} \frac{l^2}{v^2}. \quad (24-1)$$

It follows from Eq. (24-1) that for a beam of ions with equal charge  $q$  and equal initial speed  $v$ , the displacement in transit between the plates is inversely proportional to the mass  $m$  of the particular ion, all other factors ( $V_{AB}$ ,  $q$ ,  $l$ ,  $d$ , and  $v$ ) being constant for all ions by hypothesis. In continuing their flight, the "heaviest" ions will therefore hit a detector (e.g., a photographic plate) near spot  $O$  in Fig. 24-1, whereas the lighter ions may come to  $P_2$ . Conversely, from the points of incidence of the beam on the plate and from the

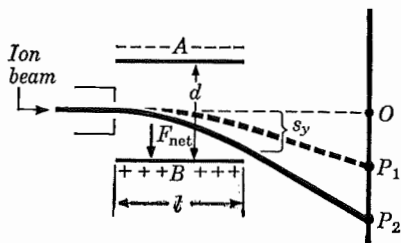


FIG. 24-1. The path of a beam of negative ions is bent in a strong electric field. For given velocity and charge, light ions fall at  $P_2$ , heavy ones at  $P_1$ , nearer  $O$ .

\* Consult the bibliography for references on mass spectrograph.

construction of the apparatus, the mass  $m$  of the ion may be determined.

In practice, it is difficult to make this method work precisely; for example, the equality of  $v$  and  $q$  for all ions is by no means easily obtained. The actual mass spectrograph is therefore designed on much more refined principles, but the final results are not different in kind from those given above.

**PROBLEM 24-1.** On the principles given, design a mass spectrograph to separate the three isotopes of uranium whose masses are approximately 238, 235, and 234 amu, where 1 amu =  $1.66 \times 10^{-24}$  gm. Assume a singly ionized beam ( $q = 1$  electronic charge) of initial speed  $v = 10^6$  cm/sec. You are free to choose  $V_{AB}$ ,  $l$ , etc. What is the separation between the points on the detector (in cm) where these three isotopes arrive? If there appear traces of the arrival of some ions just halfway between the points of incidence for  $U^{238}$  and  $U^{235}$ , what is the mass of these unknown ions? What assumptions underlie the last answer? How can we check them?

Now we are in a position to summarize some of the findings concerning isotopes. The most striking fact is probably the great variety. Some elements have only a single isotope (and, strangely, their placement numbers in the periodic table of elements, counting from hydrogen = 1 to uranium = 92, tend to be *odd*); examples are fluorine, sodium, aluminum, phosphorus. Other elements have a great number of natural isotopes; six or seven are not uncommon.\* Secondly, with the exception of some elements that have radioactive isotopes, the ratios of abundances for the isotopes are generally quite constant in a given element, no matter where it was obtained.\*\* For example, neon gas, whose isotopes were in fact the first to be separated, has three isotopes of approximate mass 20, 21, and 22 amu, in respective amounts of 90.0%, 0.27%, and 9.73% of a chemically pure sample from any natural source. Oxygen, the gas which the chemist has selected to give him a standard of relative atomic weight (16.0000 exactly), has also three isotopes,  $O^{16}$  (99.76%),  $O^{17}$  (0.04%),  $O^{18}$  (0.20%).

Thirdly, the mass of an isotope, measured relative to  $O^{16}$ , can be obtained with very great precision (six or even seven significant

\* Many more can usually be made artificially by nuclear bombardment or as fission products, but these are never stable.

\*\* Because chemical processes do not separate isotopes, and because the abundances are constant, the chemist generally does not need to concern himself at all with the concept.

figures are now usually given) and invariably is very close to a whole number. This is, of course, the justification for writing the mass number, i.e., the superscript in  $H^1$  or  $U^{238}$ , as a whole number. The mass number, which may best be defined as the true isotopic mass (in amu) rounded off to the nearest whole integer, is in fact always within less than 1% of the true isotopic mass, as Table 24-1 shows.

Table 24-1. Isotopes of a Few Elements

<i>Element</i>	<i>Natural Isotopes</i>	<i>Isotopic Mass (amu)</i>	<i>Abundance (%)</i>
Hydrogen	${}_1H^1$	1.008123	99.98
	${}_1H^2$	2.014708	0.02
Helium	${}_2He^3$	3.017016	$\sim 10^{-5}$
	${}_2He^4$	4.00388	$\sim 100$
Oxygen	${}_8O^{16}$	16.00000	99.76
	${}_8O^{17}$	17.00449	0.04
	${}_8O^{18}$	18.00369	0.20
Neon	${}_{10}Ne^{20}$	19.99872	90.0
	${}_{10}Ne^{21}$	20.99968	0.27
	${}_{10}Ne^{22}$	21.99864	9.73
Uranium	${}_{92}U^{234}$	234.1130	0.006
	${}_{92}U^{235}$	235.1156	$\sim 0.71$
	${}_{92}U^{238}$	238.1242	99.28

It is evident from the table that the isotopes of an element are generally not too far removed from one another in mass. The table also introduces a very convenient symbolism, namely, the subscript at the left of each element symbol, the *atomic number*, which refers to the place of the element in the chemist's periodic table. But at the same time we note that the physicist has not strictly adhered to the chemist's convention in the calibration of the isotopic masses. Instead of standardizing on the atomic weight of the element oxygen, i.e., of the mixture of its three isotopes in their usual proportions, as the chemist does, the physicist needs a standard of his own for isotopic mass measurements, and has chosen to call the isotopic mass of  $O^{16} \equiv 16.00000$  amu exactly. By this convention, the



weighted average of the oxygen isotopes as found in nature is 16.0044 amu on the physical scale as compared with the *chemical atomic mass* of 16.0000, by definition, on the chemical scale. The ratio between these figures is 1.00027, which is the conversion factor to use in going from one scale to the other. But actually we shall make little use of the conversion factor, and instead shall use the physical scale for discussions concerning isotopes and nuclei, and the chemical scale for problems in the chemical behavior of elements and compounds, as indeed we have tacitly done so far.

**PROBLEM 24-2.** Using the data in Table 24-1, (a) calculate the average atomic weight of the element uranium on the physical scale, and (b) convert the result to the chemical scale.

**PROBLEM 24-3.** The isotopes of ordinary argon are  ${}_{18}\text{A}^{36}$  (35.9783 amu, 0.35%),  ${}_{18}\text{A}^{38}$  (37.9745 amu, 0.08%),  ${}_{18}\text{A}^{40}$  (39.9755 amu, 99.57%). It is by no means clear at present why the abundances are distributed as they are. Calculate (approximately) what the molecular weight of argon would be if the abundances of  ${}_{18}\text{A}^{38}$  and  ${}_{18}\text{A}^{40}$  had been exchanged. Consult the periodic table, and note how this would have straightened out one of the perplexing inversions.

**PROBLEM 24-4.** Ordinary hydrogen gas consists largely of  ${}_{1}\text{H}^1$  atoms (in molecules of  $\text{H}_2$ ), but with some traces of the  ${}_{1}\text{H}^2$  isotope (called *heavy hydrogen* or *deuterium*\*) as indicated in Table 24-1. (a) Explain how the abundance of the latter could be increased by a physical process. (b) In the formation of water molecules, various isotopic combinations are possible; list them and give the molecular weight of each on the physical scale. (c) What is the molecular weight, on the chemical scale, of "heavy water," i.e., of water formed by the combination of  ${}_{1}\text{H}^2$  atoms with the usual mixture of oxygen isotopes?

**PROBLEM 24-5.** Note the doubly astonishing facts that the isotopic masses are all nearly *whole* numbers, but that they all do deviate a little from exact whole numbers. What may be the physical reasons for these two observations?

**24-2 Radioactive "half life."** After the recognition of the existence of isotopes, the next important step toward an understanding of radioactivity is the concept of *half life*. Ernest Rutherford, the man who originally noted the distinction between  $\alpha$ - and  $\beta$ -rays and who named them, suggested in 1902-1903, together with Frederick Soddy, that each of the radioactive elements has its own characteristic rate of emission of  $\alpha$ - or  $\beta$ -rays. If we make an experiment with any number  $N_0$  atoms of a specific radioactive isotope (radio-

\* From the Greek *deuteros*, second.

isotope), the number of atoms  $N_t$  left after a time interval  $t$  without having yet experienced radioactive decay follows a curve similar to Fig. 24-2. There will eventually have elapsed a time interval  $T$  when just half the original atoms ( $N_0/2$ ) survive without having yet emitted a ray. Rutherford called  $T$  the half life of the particular isotope, since an atom that has emitted no longer belongs to the type of isotopes under observation.

The principal advantage of the concept lies in the *experimental* fact that the half life can be defined at any point in the total existence of the radioisotope; no matter how old the sample is now, in another  $T$  seconds half of the present sample will still have survived. Note that this is very different for a population of, say, humans instead of radioactive atoms. If we select  $N_0$  babies, half that number may survive the 60th birthday, but of the  $N_0/2$  oldsters none is likely to celebrate a 120th birthday. However, a radioisotope sample with a half life of 60 years will still have  $N_0/4$  atoms intact after 120 years.

There are several ways of writing an expression relating  $N_0$ ,  $N_t$ ,  $T$ , and the elapsed time  $t$  of observation in a specific experiment (compare Section 12-7). We shall adopt the following form:

$$N_t = N_0 e^{-0.693t/T} \quad (24-2)$$

This is also the equation for the curve in Fig. 24-2. For a very active radioisotope, the half life  $T$  as determined by experiment is short, the exponent is large, and the curve of  $N_t$  vs.  $t$  drops rapidly to smaller values; conversely for slowly decaying substances.

Some representative experimental values for  $T$  of naturally occurring isotopes of uranium and of radium are assembled in Table 24-2. Note the variety of half-life values. There exist, for other elements, some larger and many far smaller values than those shown here; it is possible to speculate that radioactive elements, present in great quantities at the time of some hypothetical original creation of all elements, have since decayed so rapidly that no meas-

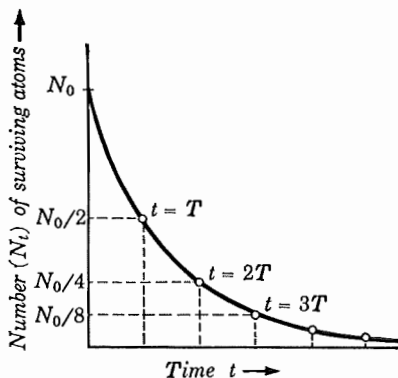


FIG. 24-2. Radioactive decay curve. After a time interval  $T$ , one-half the initial number of atoms has survived without radioactive decay.

Table 24-2. Half Lives of Some Isotopes

<i>Radioisotope</i>	<i>Half life T</i>	<i>Type of particle emitted</i>	<i>Emission energy (Mev)</i>
${}_{92}\text{U}^{234}$	$2.35 \times 10^6$ yr	$\alpha$	4.76
${}_{92}\text{U}^{235}$	$7.07 \times 10^8$ yr	$\alpha$	4.52
${}_{92}\text{U}^{238}$	$4.5 \times 10^9$ yr	$\alpha$	4.18
${}_{88}\text{Ra}^{223}$	11.2 days	$\alpha$	6.717
${}_{88}\text{Ra}^{224}$	3.64 days	$\alpha$	5.681
${}_{88}\text{Ra}^{226}$	$1.62 \times 10^3$ yr	$\alpha$	4.791
${}_{88}\text{Ra}^{228}$	6.7 yr	$\beta$	0.053

urable traces are now left. On the other hand, other elements decay so slowly that only the most refined experiments could reveal that they did not have a constant, unchanging activity.

We also realize that a chemically pure sample of natural uranium, for example, would, by the simultaneous emission of its three isotopes, complicate the data quite seriously. Sometimes, as in the case of radium, the emission of one species is different in kind from the rest ( $\beta$  instead of  $\alpha$ ), so that by the greatly larger penetrating power of one radiation over the other one can "shield out" the rest. Otherwise one can sometimes differentiate between the components by the difference of energies of similar rays, as the last column shows. The energy of  $\alpha$ -rays, given here in  $10^6$  electron volts (Mev),\* varies a little from one isotope to the other, and this may help to identify them individually, as by the difference of path length through air. Lastly, the decay curve for a mixture of isotopes will differ from that given in Fig. 24-2 for a single isotope, and a resolution of the total curve into its components can reveal the activity of the separate elements.

PROBLEM 24-6. You are given 1 gm of radium  ${}_{88}\text{Ra}^{226}$ , the long-lived and useful one among the radium isotopes. (a) On the basis of Avogadro's hypothesis, how many atoms of radium are there in the initial sample? (b) Make a plot of  $N_t$  vs. time for a period of about 8000 years. How many atoms of  ${}_{88}\text{Ra}^{226}$  are left at that time? How much total mass has been lost because of this particular decay? (c) What is the speed of the  $\alpha$ -particles of  $\text{Ra}^{226}$ ?

\* See Section 21-13: 1 electron volt = energy gained by an electron in crossing a 1 volt potential difference =  $1.603 \times 10^{-12}$  erg; this unit of energy may be used for all particles.

**PROBLEM 24-7.** For an illustration of the last point, consider Fig. 24-3, which sketches the  $N_t$  vs.  $t$  graph for three isotopes  $X^a$ ,  $X^b$ , and  $X^c$ , and for equal values of  $N_0$  in each case. If these isotopes occur in the ratios of 6:3:2, derive graphically the shape of the composite decay curve. (Note that the number of undecayed atoms  $N_t$  present at any time  $t$  compared to the initial number  $N_0$  is inferred from the relative strength of activity of the sample, e.g., from the number of particles emitted at  $t$  and counted by a Geiger counter or measured by the ionization of the gas in an electroscopes. Therefore in an actual experiment one does not plot  $N_t$  vs.  $t$ , but the equivalent information, the number of disintegrations or amount of ionization vs.  $t$ .)

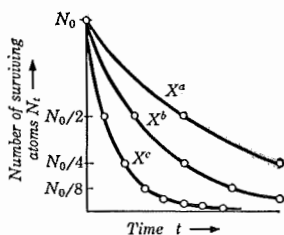


FIG. 24-3.

**PROBLEM 24-8.** Justify the statement that one can compare  $N_0$  with  $N_t$  after  $t$  sec for a sample by comparing the number of particles emitted at the start and after  $t$  sec.

**24-3 Radioactive series.** The previous two sections have seriously upset our original idea, derived from chemistry and kinetic theory, of the atom as a stable unit, uniquely given for each element. It is no wonder that the concepts of radioactive decay and of isotopes were initially difficult to assimilate. But by early 1913, both these concepts had been fitted together by several workers, principally by Soddy, in a truly beautiful theory that illuminated all the different aspects of the complex picture. One must accept, on *physical* evidence, that generally the product of radioactive decay itself is radioactive. Thus  $\text{Ra}^{226}$  decays, with the emission of an  $\alpha$ -particle, into a substance that is itself an  $\alpha$ -particle emitter, although of much shorter half life and much greater energy of emission. This in turn means a new decay product, which is again radioactive in a very individual way. But on *chemical* evidence we must also accept that each of these decay products behaves chemically in a different manner than its immediate parent and its daughter product. For instance, the metal radium ( $\text{Ra}^{226}$ ) is a product of thorium ( $\text{Th}^{230}$ ) and in turn gives rise to a rare gas, radon ( $\text{Rn}^{222}$ ). The great wealth of chemical and physical facts was summarized by recognizing that each radioisotope falls on one of three long chains, or *radioactive series*, stretching through the last rows of the periodic table. These natural series are named after an element at or near the head of each, the *uranium*, the *thorium*, and the *actinium series*.

Table 24-3. The Uranium Series\*

Energy (Mev)	4.18	0.3	0.55	4.76	4.66	4.79	5.49	6.0
Emission	$\alpha$	$\beta, \gamma$	$\beta, \gamma$	$\alpha$	$\alpha, \gamma$	$\alpha, \gamma$	$\alpha$	$\alpha$
	$^{90}\text{Th}^{234}$	$^{91}\text{Pa}^{234}$	$^{92}\text{U}^{234}$	$^{90}\text{Th}^{230}$	$^{88}\text{Ra}^{226}$	$^{86}\text{Rn}^{222}$	$^{84}\text{Po}^{218}$	
Alternate symbol	UI	UX <sub>1</sub>	UX <sub>2</sub>	UII	Io	—	RaA	
Half life	$4.5 \times 10^9$ yr	24.1 days	1.14 min	$2.4 \times 10^5$ yr	$8 \times 10^4$ yr	1620 yr	3.82 days	3.05 min
Energy (Mev)		0.65	3.15, 1.8	7.68	0.02	1.17	5.3	
Emission		$\beta, \gamma$	$\beta, \gamma$	$\alpha$	$\beta, \gamma$	$\beta$	$\alpha, \gamma$	
	$^{82}\text{Pb}^{214}$	$^{83}\text{Bi}^{214}$	$^{84}\text{Po}^{214}$	$^{82}\text{Pb}^{210}$	$^{83}\text{Bi}^{210}$	$^{84}\text{Po}^{210}$	$^{82}\text{Pb}^{206}$	
Alternate symbol	RaB	RaC	RaC'	RaD	RaE	RaF	RaG	
Half life	26.8 min	19.7 min	$1.5 \times 10^{-4}$ sec	22 yr	5 days	140 days	(stable lead)	

\* There are in this and all other series, points where a small fraction of the total radioactive decay proceeds in a branch line with different characteristics. These have been omitted for simplicity.

Table 24-3 gives the first of the chains by showing in some detail the inescapable life history of each atom of  ${}_{92}\text{U}^{238}$ . After a time  $T$  of  $4.5 \times 10^9$  years the chances are equal that it will have decayed. Having lost an  $\alpha$ -particle, the atom is now about 4 amu lighter, hence its mass number is 234. While we may name it Uranium  $X_1$ , the chemist will prove to us that the new atom is thorium, and therefore belongs in the ninetieth place in the periodic table (atomic number 90). This atom of  ${}_{90}\text{Th}^{234}$  in turn emits a  $\beta$ -ray, which does not change the mass number materially, but which, again on chemical evidence, leaves us with an atom of protactinium,  ${}_{91}\text{Pa}^{234}$ .

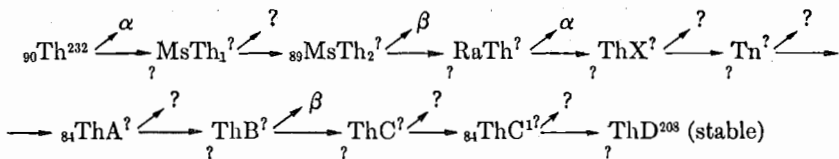
The rest of the chain shows that eventually the atom is for a time radium, then radon, and eventually stable lead ( ${}_{82}\text{Pb}^{206}$ ), although on the way it occasionally reverses its progress through the periodic table to become temporarily another isotope of an element it had previously represented. Each decay is accomplished by the emission of an  $\alpha$ - or  $\beta$ -particle of stated energy, usually associated with a  $\gamma$ -ray.

Examination of this and the other two series discloses a regularity of behavior which has the name *displacement law of radioactivity*: when an atom undergoes  $\alpha$ -decay, its mass number decreases by 4 and its atomic number decreases by 2; when an atom undergoes  $\beta$ -decay, however, its mass number does not change, and its atomic number increases by 1. Mark well that the full law was a generalization from experimental facts and, like the existence of isotopes or of radioactive decay as such, was not comprehensible in terms of any model of the atom as derived from the physical sciences of the preceding era.

PROBLEM 24-9. The actinium series begins with  ${}_{92}\text{U}^{235}$ ; each atom sends out in succession the following particles:  $\alpha$ ,  $\beta$ ,  $\alpha$ ,  $\beta$ ,  $\alpha$ ,  $\alpha$ ,  $\alpha$ ,  $\beta$ ,  $\beta$ ,  $\alpha$ . From this information and by consulting the periodic table, write out an account of the actinium series on the model of Table 24-3, complete with the necessary symbols and the figures for atomic mass and atomic number.

PROBLEM 24-10. From plutonium ( ${}_{94}\text{Pu}^{241}$ ), artificially prepared by bombarding uranium in the atomic pile, stems one of the many artificial radioactive series ("neptunium series"). The first members of the chain are  ${}_{94}\text{Pu}^{241}$ ,  ${}_{95}\text{Am}^{241}$ ,  ${}_{93}\text{Np}^{237}$ ,  ${}_{91}\text{Pa}^{233}$ ,  ${}_{92}\text{U}^{233}$ ,  ${}_{90}\text{Th}^{229}$ . Unlike all three natural series, this one does not end in a stable isotope of lead but in  ${}_{83}\text{Bi}^{209}$ . For the first six members, draw the series as in the previous problem or Table 24-3.

PROBLEM 24-11. The third natural series originates from thorium,  ${}_{90}\text{Th}^{232}$ . On the basis of the following information, replace the symbols here used and write the complete series. Indicate alternative possibilities where they can exist.



Describe briefly what experiments might have been made to determine the series as given above. (Note that this example indicates the type of interpolation that was initially often required.)

PROBLEM 24-12. Make a plot on graph paper of atomic number *vs.* mass number for the elements in the uranium series. Enter each element as a labeled dot, and draw a straight arrow from each element to its immediate decay product. This gives one of the useful and widely current representations of radioactive series.

**24-4 Rutherford's nuclear model.** We return now to the state of affairs at about 1910. The chemical and physical facts of radioactivity presented a stupendous challenge. The existence of radio elements, of  $\alpha$ -,  $\beta$ -, and  $\gamma$ -radiation, of separate and measurable half lives was well established. The idea of isotopes, arising both from the deflection of ion beams and from the first glimpses of radioactive series, was just beginning to be voiced. While the need to accommodate such findings in any one model of the atom seemed to complicate the picture beyond the power of imagination, the solution was hidden in the problem itself after all. The swift  $\alpha$ - and  $\beta$ -rays, the by-products, as it were, of radioactivity, provided exactly the tool needed to take the next important step, the means for probing the perplexing interior of the atom.

Soon after Lenard's experiment on the transmission of cathode rays through thin foils, several attempts were made to construct an atomic model which would incorporate the porosity of the atom as indicated by such experiments, as well as the presence of electrons revealed by the facts of electricity and light emission. It was variously suggested that each atom was made of myriads of massive fragments, each having a positive and a negative charge of equal amounts (Lenard's *dynamids*, 1903), or that it was a homogeneous sphere with thinly distributed mass and positive electricity, inside which there were negatively charged electrons either placed like seeds in a pumpkin or moving in circular orbits (J. J. Thomson, 1898, 1904). The attempt was made to arrange the electron positions or orbits in a stable fashion so that the mutual electric forces would not lead one to expect a dispersal or collapse of the whole structure (Thomson, Nagaoka, 1904). But the persistent and fatal shortcoming of such models was that no way existed of using them to explain



*Rutherford*

FIG. 24-4. Lord Rutherford (1871-1937).

the old puzzle of line emission spectra. Even if one wished to identify the source of emission of light with a small vibration of the electrons to either side of their equilibrium position or orbit, yet the known, complex spectra did not follow.

An entirely new basis for all speculations along such lines was provided in 1909-1911 by Ernest Rutherford (1871-1937), the New Zealander who had already shown an unmatched virtuosity in experimental physics by his work on radioactive emission in Canada, and who had moved in 1907 to Manchester University in England, where he headed a most productive research laboratory.\* Rutherford had noticed that a stream of  $\alpha$ -particles, on passing through a thin film of mica or metal, was somewhat broadened (*scattered*) by the material through which it had passed. This is interesting, for the amount of scattering might give a hint concerning the disposi-

\* Nobel prize in chemistry, 1908. For another of his many accomplishments, see Section 17.7.



tion of the atom's mass, then generally thought to be fairly evenly distributed, as in Thomson's ill-defined sphere of positive electricity. But just how Rutherford was led to the discovery that almost all the mass of the atom exists in a small central *nucleus*, has been best summarized by Rutherford himself:

“. . . I would like to use this example to show how you often stumble upon facts by accident. In the early days I had observed the scattering of  $\alpha$ -particles, and Dr. Geiger in my laboratory had examined it in detail. He found, in thin pieces of heavy metal, that the scattering was usually small, of the order of one degree. One day Geiger came to me and said, 'Don't you think that young Marsden, whom I am training in radioactive methods, ought to begin a small research?' Now I had thought that, too, so I said, 'Why not let him see if any  $\alpha$ -particles can be scattered through a large angle?' I may tell you in confidence that I did not believe that they would be, since we knew that the  $\alpha$ -particle was a very fast, massive particle, with a great deal of energy, and you could show that if the scattering was due to the accumulated effect of a number of small scatterings the chance of an  $\alpha$ -particle's being scattered backwards was very small. Then I remember two or three days later Geiger coming to me in great excitement and saying, 'We have been able to get some of the  $\alpha$ -particles coming backwards . . .' It was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you. On consideration, I realized that this scattering backwards must be the result of a single collision, and when I made calculations I saw that it was impossible to get anything of that order of magnitude unless you took a system in which the greater part of the mass of the atom was concentrated in a minute nucleus. It was then that I had the idea of an atom with a minute massive center carrying a charge. I worked out mathematically what laws the scattering should obey, and I found that the number of particles scattered through a given angle should be proportional to the thickness of the scattering foil, the square of the nuclear charge, and inversely proportional to the fourth power of the velocity. These deductions were later verified by Geiger and Marsden in a series of beautiful experiments."\*

\* From the essay (1936) "The Development of the Theory of Atomic Structure," by Lord Rutherford, in *Background to Modern Science*, New York: Macmillan Co. (1940).

This, then, was the birth of the nuclear atom—a conception almost immediately accepted into science despite the initial uncertainties of detail. Because it gave the picture of a central concentration of a (positively charged) mass about which the light electrons can swarm at relatively great distances, the whole problem of atomic theory was immediately simplified. For now one could think in separate terms about the very different phenomena for which the atom had to be made responsible; the extranuclear electrons would be concerned with chemical behavior, electricity, the emission of light, etc., whereas the nucleus would serve for the phenomena of radioactivity, the explanation of isotopes, and the like.

Now we turn to the quantitative details of the work of Rutherford and his co-workers, H. Geiger and E. Marsden. The projectiles used were  $\alpha$ -rays with speeds of about  $2 \times 10^9$  cm/sec, fired in a narrow beam against metallic films, e.g., a gold foil only  $4 \times 10^{-5}$  cm thick. On the basis of the approximate atomic dimensions deduced from kinetic theory, one may expect such a film to represent about  $10^3$  layers of atoms. Most of the  $\alpha$ -particles do go right through the foil without noticeable deviation, and therefore we have additional evidence for the proposition that the atoms are *not* "solid." They must contain electrons, but these can hardly be held responsible for changing the path of those  $\alpha$ -particles that are scattered so drastically as to come almost straight back. At each encounter with an electron, the  $\alpha$ -particle can experience only a very minute change in momentum because of the great disparity between the respective masses. The  $\alpha$ -particle has a mass of  $m_\alpha \doteq 6.85 \times 10^{-24}$  gm, whereas the mass of the electron,  $m_e$ , is only  $9.11 \times 10^{-28}$  gm, i.e., about 1/7500 as large; even if we assume a head-on collision, the forward component of the velocity of the  $\alpha$ -particle changes only very little each time—to be precise, about 0.01%. The probability is vanishingly small that an  $\alpha$ -particle, in thousands of successive encounters with electrons, will experience small deviations from its original direction that are so predominantly in one sense that the net result is an actual turning back. The alternative explanation of the observed scattering is, therefore, that an  $\alpha$ -particle may meet a single object in the foil that is heavy enough to turn it back in one encounter, just as a bullet will ricochet back from a boulder but not from a loose pile of sand particles.

PROBLEM 24-13. An  $\alpha$ -particle of mass  $m_\alpha$  with initial speed  $v_{\alpha 1}$  approaches an electron with mass  $m_e$ , the latter initially at rest. After the head-on (perfectly elastic) encounter, the  $\alpha$ -particle continues with speed  $v_{\alpha 2}$

and the electron moves with speed  $v_e$ . (a) Write the two equations for conservation of momentum and of energy for this event, and prove that

$$v_e = \left( \frac{2 m_\alpha}{m_\alpha + m_e} \right) v_{\alpha 1}, \quad v_{\alpha 2} = \left( \frac{m_\alpha - m_e}{m_\alpha + m_e} \right) v_{\alpha 1}.$$

(b) Calculate the actual change in speed and in momentum for an  $\alpha$ -particle in one encounter (in percentage).

In the paragraphs above we have assumed that an  $\alpha$ -particle will interact with an electron as though this were a mechanical collision between marbles, but of course we should consider now that both are oppositely charged particles. The mechanism of interaction is then primarily not an elastic collision but rather a mutual attraction, the same effect by which the sun causes a comet that is passing close by to turn around after half circling the sun. To be precise, the comet's path will be hyperbolic if the comet does not remain in the solar system but escapes again from the gravitational attraction between it and the sun. But as for the momentum and energy changes, those are the same no matter whether the mechanism of interaction is elastic or electric. The fact of occasional large-angle scattering of  $\alpha$ -particles is not explainable by multiple encounters with electrons even if we consider that the force of interaction is electric attraction.

On the other hand, visualize the hypothetical behavior of an  $\alpha$ -particle if it encountered a small heavy nucleus of positively charged mass. Figure 24-5(a) is based on one in Rutherford's pub-

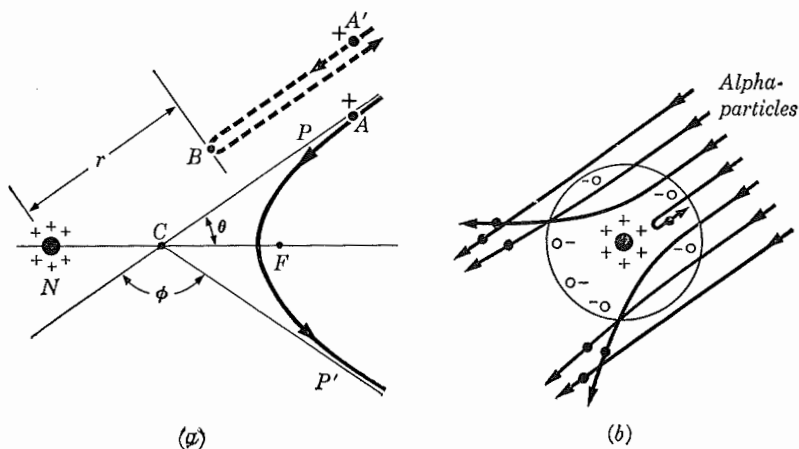


FIG. 24-5.

lication "The Scattering of  $\alpha$ - and  $\beta$ -Particles by Matter and the Structure of the Atom," in the scientific British journal, *Philosophical Magazine and Journal of Science* ("Phil. Mag.") in 1911, which may be said to have laid the foundation for the modern theory of atomic structure. The  $\alpha$ -particle at  $A$  is approaching a nucleus at  $N$ , is repelled by the (Coulomb's law) force between the positively charged bodies, and follows a path  $PP'$  that must be a hyperbola.\*

Now one may inquire how large the angle of deviation or scattering from the original direction should be for a parallel beam of  $\alpha$ -particles approaching a layer of distributed nuclei. Particle  $A'$ , for example, is heading directly toward  $N$ , approaches only up to point  $B$ , and then returns directly back; most other particles do not pass so close to any nuclei and so do not experience such serious deflections [Fig. 24-5(b)].

Rutherford considered the effect on  $\alpha$ -particle scattering of all important factors such as charge on the nuclei, speed of incident particle  $v_\alpha$ , and thickness of film  $t$ , and derived the following predictions. If scattering occurs by single encounters with small positive nuclei, the fraction of  $\alpha$ -particles scattered by angle  $\phi$  should be directly proportional to (a)  $[\sin(\phi/2)]^{-4}$ , (b) the foil thickness  $t$ , (c) the magnitude of the nuclear charge, and (d)  $(v_\alpha)^{-4}$ . Predictions (a), (b), and (d) could be checked quickly; in fact, the previous year Geiger had published scattering measurements which were most favorable, and a more complete and fully convincing check was given in 1913 by Geiger and Marsden.

Figure 24-6 represents schematically the apparatus used by Geiger and Marsden.  $B$  is a lead box from which an  $\alpha$ -emitter

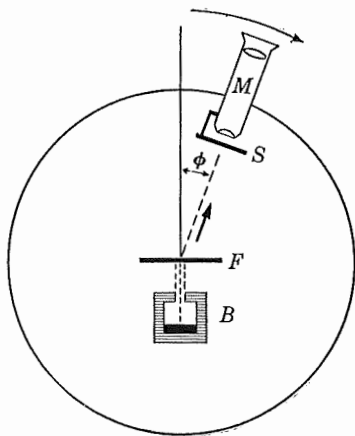


FIG. 24-6. Diagram of apparatus for verifying Rutherford's theory of alpha-particle scattering.

\* On the basis of Newton's *Principia* it had long been known that if a body  $A$  moves toward another,  $N$ , against a force of repulsion which is inversely proportional to the square of the distance between them, as is the case with these positively charged particles, then the path taken by  $A$  is the same as if, instead of  $N$ , there were a similar but oppositely charged body at  $F$ , the focus of the curve. This defines a hyperbola with center at  $C$ .

(radon) sends out a beam through a small diaphragm; the foil  $F$  scatters the rays in varying degrees through different angles  $\phi$ . The number of scattered particles at each angle is found by letting them fall upon a zinc-sulphide screen  $S$ , where each incidence causes a momentary flash of light (scintillation) observable and countable through the microscope  $M$ ;  $S$  is fixed to  $M$  and can be rotated in this plane up to  $\phi = 150^\circ$ .

PROBLEM 24-14. If about 100  $\alpha$ -particles of equal speed were observed to penetrate a given foil each minute without any scattering, plot on the basis of Rutherford's prediction a graph of number of scintillations observed per minute *vs.* angle  $\phi$  of scattering.

While the observation at different angles  $\phi$ , with foils of different  $t$ , and with  $\alpha$ -particle beams of different  $v_\alpha$ , confirmed three of Rutherford's four predictions, there appeared no direct way of checking on the effect of the nuclear charge on scattering. However, one could turn the argument around, assume that the experiments had sufficiently proved the general supposition, and then calculate the nuclear charge from the data. This was done for carbon (in a sheet of paraffin wax), and for metal foils from aluminum to gold, and it turned out that for these materials the nuclear charge, expressed in multiples of the numerical electronic charge  $e$  ( $4.8 \times 10^{-10}$  statcoul), was roughly  $\frac{1}{2}$  (at. wt.)  $\times e$ . This would be 6  $e$  for carbon, 13  $e$  or 14  $e$  for aluminium, 78  $e$  or 79  $e$  for gold, and so on.

These experiments therefore provided simultaneously a vindication of Rutherford's hypothesis, and an important piece of new information, for if the nucleus carries a charge equal to ( $\frac{1}{2}$  at. wt.) units of electronic charges, there must be equally many electrons around it to provide for the electric neutrality of the whole atom. In fact, previous experiments on the scattering of x-rays and  $\beta$ -particles had given the number of electrons about equal to ( $\frac{1}{2}$  at. wt.) also. It was pointed out soon afterwards by A. van den Broek in 1913 that the important comparison to make was not between nuclear charge and half of the atomic weight, but between the former and *atomic number* ( $Z$ ). It is an astonishing fact that for most substances,  $Z = \frac{1}{2}$  at. wt. within narrow limits; e.g.,  $Z$  for carbon is 6, the at. wt. is 12.010. Thus the experimentally founded suggestion that the number of positive charges on the nucleus or, equally well, of electrons around the nucleus, be simply equal to the atomic number  $Z$  at once made the picture of the atom more precise and complete. A hydrogen atom ( $Z = 1$ ) in its neutral state has

one electron moving around its positively charged nucleus, a helium atom ( $Z = 2$ ) has 2 electrons, and so on up to uranium ( $Z = 92$ ) with 92 extranuclear electrons. This simple scheme was made more plausible when in ion beams in the mass spectrograph there were found singly ionized hydrogen ( $H^+$ ) and even doubly ionized helium ( $He^{++}$ ), but never  $H^{++}$  or  $He^{+++}$ , evidently because H atoms only had a single electron to lose, and He atoms only two.

A second important consequence from scattering experiments concerned the size of the nucleus, which Rutherford estimated by an ingenious argument. Since the rest of his deductions depended on the validity of Coulomb's law of electric forces between the nucleus and  $\alpha$ -particle, and since these deductions (as of the variation of scattering with angle) had been borne out, we may assume Coulomb's law to hold during the scattering process. Now surely it would not hold if the projectile were to come so close to the nucleus as to deform it elastically or touch it in any way. Therefore the closest distance of mutual approach between the centers of the target nucleus and the  $\alpha$ -particle must be less than the sum of their radii. The actual distance of nearest approach can, however, be calculated, for in such a case the  $\alpha$ -particle must be heading straight toward the nucleus [see path  $A'B$ , Fig. 24-5(a)] and will penetrate to point  $B$  at a distance  $NB \equiv r$  from the nucleus, where the kinetic energy, by the work against mutual repulsion, is all converted to potential energy; after this the projectile turns back like a stone at the top of its trajectory. At  $B$ , the point of momentary rest,  $KE_{\text{init}} = PE$  or, numerically,

$$\frac{1}{2} m_{\alpha} v_{\alpha}^2 = \frac{(Ze) \cdot (2e)}{r},$$

where  $v_{\alpha}$  is the initial speed of the  $\alpha$ -rays,  $(2e)$  its charge, and  $(Ze)$  the charge on the target nucleus  $N$ . [Here we have assumed that the potential energy of the  $\alpha$ -particle at distance  $r$  is given by Eq. (21-6) for point charges.]

For a gold foil bombarded with  $\alpha$ -particles of about  $2 \times 10^9$  cm/sec speed, the distance  $r$  can be calculated to be about  $3 \times 10^{-12}$  cm (see problem 24-15). Compared to the over-all size of the atom ( $\sim 10^{-8}$  cm), the nuclear size must therefore be of the order of  $10^{-12}$  or  $10^{-13}$  cm, i.e., less than 1/10,000 as large! The mass of electrons being almost negligible compared with that of even the lightest atom as a whole, most of the mass of the atom is thus concentrated within a radius that bears to the total atomic radius a relation similar to that between the sun's radius and the radius of the total solar system

measured to Pluto's orbit. One may assume that it is the presence of moving electrons around the nucleus at the relatively large distance of  $10^{-8}$  cm which determines the "size" of the atom. Evidently this scheme presents us with an atom that is mostly empty space, as indeed it must be to account for the relative ease of penetration of  $\alpha$ - and  $\beta$ -radiation through thousands of atoms, as in a metallic foil or in the gas of a cloud chamber.

**PROBLEM 24-15.** Determine the distance  $r$  of nearest approach between an  $\alpha$ -particle with  $v_\alpha = 2 \times 10^9$  cm/sec, and a nucleus of a gold atom. (This calculation was made by Rutherford in his 1911 paper, but he considered  $Z \doteq \frac{1}{2}$  at. wt.)

Immediately, the developing atom raises three issues: (a) What is the structure of the nucleus? (b) What keeps the nucleus in general from exploding by the mutual repulsion of its components if it is made of densely packed charges of equal sign? (c) How are the electrons arranged around the nucleus?

To the first of these questions, Rutherford initially could do no more than to suggest that at least in some elements the charge and mass of the nucleus can be explained by assuming a grouping of a number of helium nuclei. "This may be only a coincidence, but it is certainly suggestive in view of the expulsion of helium atoms [ $\alpha$ -particle] carrying two unit charges from radioactive matter."

**PROBLEM 24-16.** According to this suggestion, how many helium nuclei would make up one nucleus of carbon, of nitrogen, of oxygen? Does the scheme work for all elements in the periodic table?

While therefore the first question was actually unsolvable at the time, the second one concerning nuclear stability was even more puzzling, and had to be dismissed by Rutherford as follows: "The question of the stability of the atoms proposed need not be considered at this stage, for this will obviously depend upon the minute structure of the atom, and on the motion of the constituent charged parts." Rutherford's later work on artificial transmutations opened the way to a better answer to both problems, as we shall see at once.

Finally, the third question raised above was destined to be solved soon by Bohr, another co-worker of Rutherford's. At the time, Rutherford simply noted that the Japanese physicist Nagaoka had speculated in 1904 on a model of the atom in which a central nucleus was pictured as surrounded by electrical charges revolving in rings similar to those seen around Saturn.

PROBLEM 24-17. If in a gold foil the atoms are so close as to touch one another, with a distance of about  $3 \times 10^{-8}$  cm between their nuclei, and if the radii of the nuclei of the atoms and of the  $\alpha$ -particles are for this rough calculation each assumed to be about  $3 \times 10^{-12}$  cm, calculate the probability of an  $\alpha$ -particle approaching a nucleus virtually head-on in any one layer of metal atoms. (*Hint:* Divide the diameter of a gold atom into small sections, each equal to the diameter of the  $\alpha$ -particle; then assume that the chances for the  $\alpha$ -particle to pass through any section are evenly distributed.)

**24-5 Moseley's x-ray spectra.** In the year 1913, when Geiger and Marsden's scattering experiments completely verified the main features of the theory of the nuclear atom, there came yet another profound contribution from Rutherford's laboratory. There the young physicist H. G. J. Moseley (1887-1915) (Fig. 24-7) had been investigating the wavelengths of x-rays emitted when a stream



FIG. 24-7. H. G. J. Moseley (1887-1915). (From the collection of Professor George Sarton.)



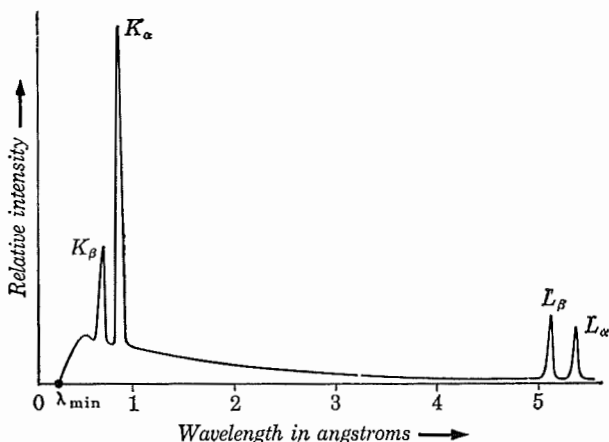


FIG. 24-8.

of high-speed electrons (cathode rays) impinges on metallic targets (see Fig. 23-9). A plot of the intensity of x-rays *vs.* wavelength gives a graph rather like Fig. 24-8. It reminds one somewhat of a continuous emission spectrum of light from a hot solid, but with a line spectrum superimposed on it. Indeed, we might analyze the graph in two parts. The continuous spectrum of so-called *general radiation*, which serves as “background” for the sharply defined spikes, is generated by the conversion of part of the kinetic energy of cathode rays into photons as they pass through the atoms of the target [see Eq. (23-10)]. Those electrons that give up all their energy in one interaction with the target atoms give rise to the highest frequency x-ray photons by the relation  $h\nu = \frac{1}{2} m_e v^2$  [Eq. (23-7)], but as  $\lambda = c/\nu$ , the shortest wavelength in Fig. 24-8 is  $\lambda_{\min} = hc/\frac{1}{2} m_e v^2$  and is well defined for a given electron beam.

**PROBLEM 24-18.** (a) What voltage must be applied to the x-ray tube to obtain a continuous spectrum with  $\lambda_{\min} = 0.5 \text{ \AA}$ ? (Recall that, in consistent units,  $\frac{1}{2} m_e v^2 = Ve$ , where  $V =$  potential difference between electrodes,  $e =$  charge on electron.) (b) How would you use these relations to measure the ratio  $h/e$  with precision? (c) Show that  $\lambda_{\min} = 12,400/V$ , where  $V$  is in volts.

Moseley, however, was not concerned with the continuous spectrum (which was not fully understood at the time). His work concentrated on the other part of the total x-ray spectrum, the “spikes” or lines or *characteristic radiation*. For each of his targets, the re-

sult was similar to Fig. 24-8; there were several such lines, usually in two groups named *K* and *L*, although later, better equipment brought out more groups (*M*, *N*, . . .) and, in each group, usually many more lines. Nevertheless, even the initial data revealed an astonishing simplicity. The lines in the characteristic x-ray spectrum progressively shifted to shorter wavelengths as the *atomic number* of the target material increased. Table 24-4 compares the prominent  $K_{\alpha}$  and  $K_{\beta}$  lines of five neighboring metallic elements in the periodic table; in each row, the wavelength of  $K_{\alpha}$  or  $K_{\beta}$  is shorter than in the preceding row.

Table 24-4.

Wavelengths of Prominent Characteristic K Lines (in  $10^{-8}$  cm)

<i>Element</i>	<i>At. number</i>	<i>At. wt.</i>	$\lambda$ of $K_{\alpha}$	$\lambda$ of $K_{\beta}$
Manganese (Mn)	25	54.93	2.10	1.91
Iron (Fe)	26	55.85	1.93	1.75
Cobalt (Co)	27	58.94	1.79	1.62
Nickel (Ni)	28	58.69	1.66	1.50
Copper (Cu)	29	63.54	1.54	1.39

At once, several conclusions could be drawn from this discovery of regularity. (a) From his initial data on the characteristic radiation for nine elements from calcium to copper, and from his further work in 1914, which embraced elements between aluminum ( $Z = 13$ ) and gold ( $Z = 79$ ), Moseley found that the frequency  $\nu$  of the  $K_{\alpha}$  line could generally be given by an empirical equation that may be written

$$\nu(\text{sec}^{-1}) = 2.48 \times 10^{15} (Z - 1)^2. \quad (24-3)$$

(An analogous equation was given for  $L_{\alpha}$  lines.) Such an equation has immediate uses. For example, although the metal scandium ( $Z = 21$ ) was not then available as a target material, the frequency of  $K_{\alpha}$  could be calculated with considerable confidence. When that line was later measured, the agreement was, in fact, found to be good. This is reminiscent of the story connected with Balmer's formula for the ultraviolet portion of the hydrogen spectrum, and indeed, both formulas must find a similar type of physical explanation in the final theory of the atom.

**PROBLEM 24-19.** The prominent  $K_{\alpha}$  line of scandium has a wavelength of  $3.03 \times 10^{-8}$  cm. Check to what extent Moseley's empirical formula holds for this line.

(b) The predictive value of Moseley's formula is, of course, most gratifying. So much confidence could later be placed in it that the identification of a newly discovered element depended on its having characteristic x-ray lines that fitted properly between those of its immediate neighbors in the periodic table, and conversely, wherever a great gap occurred between the wavelengths of  $K_{\alpha}$  for neighboring elements, there, and only there, could one suspect the presence of a yet unknown element.

Yet, Moseley had a rather different point in mind in originally undertaking his research, a problem which he brilliantly illuminated. It must be remembered that in Moseley's time the concept of *atomic number* was not yet of importance in physics; the idea was in the air that it corresponded to the number of positive charges on the nucleus, as Rutherford said later, but it had not been announced. In fact, it was Moseley who seems to have first used the term atomic number explicitly. It requires only one glance at Table 24-4 to see that the progressive regularity in the  $K$  lines of elements is dependent on the atomic number and not on the atomic weight. Cobalt and nickel represent an inversion in the periodic table, yet the  $K$  lines follow regularly. (The same point was established for the other inversions.)

Here is comfort for the chemist who was forced to make inversions on the basis of chemical behavior and macroscopic physical properties, and here is also proof that whatever be the cause of the characteristic x-ray spectrum, it is connected with the atomic number, i.e., with the number of positive charges on the nucleus. This is quite reasonable, for we cannot help but guess, pending a full explanation, that x-rays are generated when the penetrating cathode ray sets the electrons near the nucleus into such motion as will cause the emission of high-energy photons. But the behavior of an electron near the nucleus surely must be governed, above all, by the field strength in that region owing to the charges on the nucleus. Moreover, as Moseley pointed out, "The very close similarity between the x-ray spectra of the different elements shows that these radiations originate inside the atom, and have no direct connection with the complicated light spectra and chemical properties governed by the structure of its surface."

(c) That x-rays are purely the property of atoms was also made clear by noting that if the target in the x-ray tube were made of brass, an alloy of copper and zinc, the undisturbed characteristic radiations of both Cu and Zn atoms were observed simultaneously.

To summarize the central points in Moseley's own words, "We have here a proof that there is in the atom a fundamental quantity, which increases by regular steps as we pass from one element to the next. This quantity can only be the charge on the central positive nucleus, of the existence of which we already have definite proof." A great deal of research soon made it quite certain that this charge, measured in units of electronic charge but with a positive sign, was in every case exactly equal to the atomic number. Moreover, here we have at last been given one property that changes in a regular way from one element in the periodic table to the next. The chemical properties, such as valence, and the physical behavior, such as melting points, do not vary in continuous gradation but in definite periods, with sudden breaks between one and the next. Even the atomic weight progresses unevenly, as the inversions prove. Some other features, such as line-emission spectra, establish no simple relation whatever between neighboring elements. Only the progressively increasing charge on the nucleus gives us a truly secure foundation for the arrangement of elements in a unique order.

**24-6 Further concepts of nuclear structure.** Because this chapter is to be in the main an account of the theory of nuclear structure, we may briefly deviate from the historical development to summarize a few of the more recent concepts of nuclear physics. The emission from natural radioactive elements of  $\alpha$ - and  $\beta$ -particles suggested naturally that a model of the nucleus start with these building blocks. But the very first element, hydrogen, offered difficulty, having but a single positive nuclear charge and about one-fourth the mass of an  $\alpha$ -particle.\* The nucleus of the usual atom of hydrogen was accessible for observation, either in the form of  $H^+$ , the ionized atom whose single electron had been removed in the violence of an electric discharge, or as the result of bombardment of hydrogen gas with high-speed  $\alpha$ -particles, where occasionally a  ${}_1H^1$  nucleus would be seen to speed through the cloud chamber or to a scintillation screen

\* The heavy isotope of natural hydrogen, which has about half the mass of an  $\alpha$ -particle, was not known until 1931, and of course cannot be explained as a combination of an  $\alpha$ - and a  $\beta$ -particle.

after having experienced a collision. The name *proton* was generally adopted for the particle in 1920. The previous year, Rutherford had reported that such particles appeared to be emitted when nitrogen is bombarded by  $\alpha$ -particles (as was discussed in Section 17-7): “. . . we must conclude that the nitrogen atom is disintegrated under the intense force developed in a close collision with a swift  $\alpha$ -particle, and that the hydrogen atom which is liberated formed a constituent part of the nitrogen nucleus.” He added prophetically, “The results on the whole suggest that, if  $\alpha$ -particles—or similar projectiles—of still greater energy were available for experiment, we might expect to break down the nuclear structure of many of the lighter atoms.”

Now the problem of nuclear structure seemed much nearer to solution. Each nucleus carries a number of protons equal to its atomic number  $Z$ . Each proton has one unit charge and a mass of about 1 amu, so that this scheme so far accounts for the proper positive charge but, in general, for only half the mass; we may therefore imagine the nucleus to contain sufficient neutral pairs of protons and (negative) electrons to make up the difference between the observed atomic weight and the atomic number. Thus a helium nucleus would be composed of 4 protons and 2 nuclear electrons, to give a charge of  $+2$  and a mass of about 4 (the electrons having only  $1/1836$  of the mass of a proton, therefore not adding significantly to the total nuclear mass). Similarly, a nucleus of  ${}_{92}\text{U}^{238}$  would contain 92 un-neutralized protons and  $(238 - 92)$  neutral proton-electron pairs.

This hypothesis has two appealing features. It places in the nucleus, electrons that are known to come from there in radioactive decay as  $\beta$ -radiation, and it allows us to understand at last why an element, after  $\alpha$ - or  $\beta$ -decay, is found in a different position in the periodic table. Since the atomic number is given by the number of positive nuclear charges, the loss of two positive units by  $\alpha$ -emission should move the element down two places, and the gain of one positive unit through the loss of a (negative) nuclear electron in  $\beta$ -emission should move the element up by one place—all just as observed.

The hypothetical combination proton-electron was given the appropriate name *neutron* in the early 1920's, and a wide search was made to discover it experimentally in a free state, e.g., as a product of inelastic collision of fast  $\alpha$ -particles with other nuclei. None of the experiments planned specifically for this purpose succeeded, but in 1930 it was reported that energetic  $\alpha$ -particles, when stopped by

light elements (lithium, beryllium, boron) could create a radiation so penetrating that they were considered to be high-energy  $\gamma$ -rays. Indeed, the energies obtained were greater than of any then known  $\gamma$ -radiation. In a historic experiment, the French physicists Frédéric Joliot and his wife Irène Curie (daughter of the discoverers of radium) noted in 1932 that this new radiation could in turn eject protons of very high speed from hydrogen-containing materials such as paraffin wax. Immediately after the publication of this result, the British physicist James Chadwick, who had been an assistant of Rutherford's, showed that the protons could not have been ejected by a  $\gamma$ -radiation unless the latter not only differed in behavior from all other known electromagnetic radiations but also, if one wished to account for the great speed given to the protons, the  $\gamma$ -radiation would somehow have to leave the parent atoms with an energy 10 times larger than had been given up by the  $\alpha$ -particles that initially created it. In his paper "The Existence of a Neutron," Chadwick then goes on to prove that "the experimental results were very difficult to explain on the hypothesis that the beryllium radiation [the rays created when  $\alpha$ -particles fall on Be] was a quantum [ $\gamma$ , photon] radiation, but followed immediately if it were supposed that the radiation consisted of particles of mass nearly equal to that of a proton and with no net charge, or neutrons." Having no charge, these neutrons are non-ionizing, therefore do not lose their energy easily in passing through matter, can penetrate thick sheets of lead, and can approach a charged nucleus directly without experiencing a force of repulsion. In a head-on collision with a hydrogen nucleus, of approximately equal mass, the neutron's energy would be virtually completely transmitted to the proton.

At once it became clear that the neutron, together with the proton, would account completely for the nuclear structure of all elements. But one no longer thinks of the neutron as a particle composed of 1 proton and 1 electron; instead, the neutron is regarded as a fundamental particle that may, under certain conditions, decay into a proton, an electron ( $\beta$ -ray) and, presumably, a neutrino (see Section 17-7).

Figure 24-9 represents schematically the structure of several nuclei ( $n$  = neutron,  $p$  = proton). For  ${}_1\text{H}^2$ , the "heavy" isotope of hydrogen, or *deuterium*, one neutron and one proton are required, a combination referred to as a deuteron. The helium nucleus or  $\alpha$ -particle comprises two neutrons and two protons. Similarly for the other nuclei.

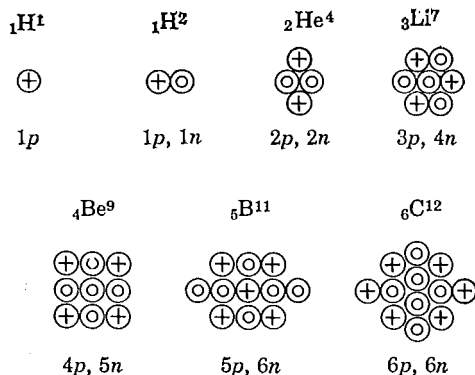
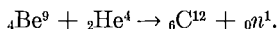


FIG. 24-9. Purely pictorial representation of some nuclei;  $n$  = neutrons,  $p$  = protons.

**PROBLEM 24-20.** (a) Draw similar representations for  ${}^8_8\text{O}^{16}$ ,  ${}^8_8\text{O}^{17}$ ,  ${}^8_8\text{O}^{18}$ ,  ${}^{15}_{15}\text{P}^{31}$ . (b) How many protons and neutrons are there in each of the natural isotopes of uranium?

**PROBLEM 24-21.** The nuclear reaction in which Chadwick discovered the neutron is a transmutation that may be written as follows



The rest masses of neutral isotopes are  ${}^4_4\text{Be}^9 = 9.01506$  amu,  ${}^2_2\text{He}^4 = 4.00388$ ,  ${}^6_6\text{C}^{12} = 12.00386$ ; mass of neutron ( ${}^1_0n^1$ ) = 1.00896 amu; also, the energy of the bombarding  ${}^2_2\text{He}^4$  nuclei, obtained from polonium, is known to be about  $5.3 \times 10^6$  electron volts. Calculate the energy and speed of the emerging neutrons. (Recall that  $1 \text{ amu} \doteq 1.66 \times 10^{-24} \text{ gm}$ ,  $1 \text{ ev} \doteq 1.60 \times 10^{-12} \text{ erg}$ . Clearly express any assumptions you may make in your calculations.)

We may conclude the account of the nuclear atom with a brief summary of some typical concepts that are useful in contemporary research. Most important is probably the question of the stability of the nucleus, the problem that Rutherford originally wisely bypassed. Mechanisms have been invented that account adequately for the main facts of stability and for the general facts of natural and induced radioactive decay, although by no means in the framework of the classical physics which we have largely relied on so far in describing the nuclear model. To us one most intriguing point is that the neutron and proton, having about equal mass, very close to 1 amu, necessarily create nuclei that have almost integral masses on the amu scale.

Thus we are led back to Prout's hypothesis, although in a man-

ner he did not envisage at all, and we now understand the previously most mysterious fact that all isotopic masses are nearly whole integers. On the other hand, we are quite ready to expect that the isotopic masses will not all be *exactly* integral multiples of  ${}_1\text{H}^1$ , not only because the neutron has a slightly larger mass than the proton (1.00896 compared with 1.00758), but rather because of the much more important effect on the rest mass owing to different potential energies among the particles within the different nuclei. As has been discussed (Section 17-7), the nucleus of a helium atom has a smaller mass (4.0028 amu) than the sum of the individual particles in their free states (2 protons =  $2 \times 1.00758$ , 2 neutrons =  $2 \times 1.00896$ , together 4.0331 amu), a difference  $\Delta m$  of about 0.0303 amu that corresponds to the energy [ $\Delta E = (\Delta m)c^2$ , about  $4.5 \times 10^{-5}$  erg or 28 Mev] that must be supplied to a helium nucleus to break it into its constituents. This energy, called *binding energy* of the nucleus, although perhaps the term *unbinding energy* is more descriptive, can be similarly calculated for the other nuclei, and summarizes in a convenient way the energy relationship in nuclear reactions. The corresponding difference of mass,  $\Delta m$  itself, is sometimes called the *mass defect*.

PROBLEM 24-22. Table 24-5 gives the mass of a few stable elements for the most abundant isotope of each. Calculate the binding energy of each (recall that 1 amu  $\doteq 1.49 \times 10^{-3}$  erg), and plot a graph of mass number (abscissa) *vs.* binding energy per nuclear particle. This is the usual way of presenting this information for use in problems of nuclear physics. (*Note:* The data in Table 24-5 are isotopic masses for the *neutral atom*. To obtain *nuclear masses*, subtract the mass of the corresponding extranuclear electrons, or 0.00055 amu per electron.)

Table 24-5

<i>Most abundant isotope</i>	<i>Atomic mass (amu)</i>
${}_1\text{H}^1$	1.00813
${}_4\text{Be}^9$	9.01506
${}_7\text{N}^{14}$	14.00751
${}_{10}\text{Ne}^{20}$	19.99872
${}_{15}\text{P}^{31}$	30.9839
${}_{26}\text{Fe}^{56}$	55.9534
${}_{41}\text{Cb}^{93}$	92.926
${}_{57}\text{La}^{139}$	138.955
${}_{82}\text{Pb}^{208}$	208.057



PROBLEM 24-23. The *packing fraction* of a nucleus is defined as follows:

$$\text{Packing fraction} = \frac{\text{Isotopic weight} - \text{Mass number}}{\text{Mass number}}$$

It is a useful concept for comparing the stability of nuclei, for evidently a large negative packing fraction implies a large mass defect and hence a large amount of energy for breaking up the structure. Using the data of Table 24-5, plot a graph of packing fraction *vs.* mass number. What is the physical significance of the minimum in the curve?

Now that we have reached a model of the nucleus that is actually as close to pictorialization as is permissible in a subject so completely given over to mathematical rather than physical models, we must turn to the final questions: What is the structure of the swarm of electrons that we have tentatively postulated around the nucleus? What are the laws governing their behavior, and to what extent can they account for our remaining puzzles of chemistry and physics, e.g., valences and emission spectra? When these questions have been answered, we shall indeed be in possession of the key to a unified understanding of the major questions of physical science.

### Further References

R. T. Beyer (ed.) *Foundations of Nuclear Physics*. New York: Dover Publications (1949). Reprints of original papers by Rutherford and by Chadwick.

H. Buckley, *op. cit.* (Ch. 15). Chapters XI and XII.

T. W. Chalmers, *op. cit.* (Ch. 17). Chapter XII on isotopes.

A. S. Eve, *Rutherford*. New York: Macmillan (1939). A fine biography.

S. Glasstone, *Sourcebook on Atomic Energy*. New York: D. Van Nostrand Co. (1950). A very useful reference on every topic of nuclear physics.

W. F. Magie, *op. cit.* (Ch. 2), pp. 600-616. Excerpts from Roentgen, Becquerel and the Curies. Also J. W. Knedler, *op. cit.*, pp. 571-595 (Curie).

C. Møller and E. Rasmussen, *The World and the Atom*. London: George Allen and Unwin (1940). Chapters I and III. An excellent account of the development of nuclear physics.

O. Oldenberg, *op. cit.* (Ch. 23), particularly Chapters 8 and 20 on isotopes and measurements of their mass.

H. Semat, *Introduction to Atomic Physics*. New York: Rinehart and Co. (1946). Chapters 8 and 9 on Disintegration of Nuclei and on Nuclear Energy.

H. D. Smyth, *A General Account of the Development of Methods of Using*

*Atomic Energy for Military Purposes* etc. Washington: U. S. Government Printing Office (1945). The famous report—generally very readable.

W. C. Dampier Whetham and M. Dampier Whetham, *Cambridge Readings in the Literature of Science*. Cambridge: Cambridge University Press (1924). Pp. 132-167 give extracts from Thomson, Moseley, and Rutherford.

## BOHR'S MODEL OF THE ATOM

**25-1 Introduction.** This final chapter is an account of the original work in 1913 of a young Danish physicist, and its consequences for the physical sciences. Niels Bohr (page 471), born in 1885, the year of Balmer's publication concerning the hydrogen spectrum, had just received his Ph.D. degree from the University of Copenhagen in 1911 when he joined J. J. Thomson in Cambridge as visiting researcher. After a few months he left to join Rutherford's group at Manchester. The work at that laboratory was just bringing proof of the theory of nuclear atoms; at the same time, the rest of the atom, the configuration of the electrons about the nucleus, was still a puzzle. There were only the more or less general suggestions that the electrons moved about the nucleus in closed orbits, a scheme rather analogous to the planetary system. Indeed, Thomson recently had been able to show that the system would be stable if one assumed the electrons to be arranged in rings, with 2 electrons on the innermost ring, up to 11 in a second ring, and so on.

But the hypotheses had two fatal weaknesses. First, although they provided electrons to account for the *presence* of the phenomena of electric currents, ionization, photoelectricity, the neutrality of the ordinary atom, x-ray scattering, and so forth, they were unable to produce quantitative explanations for the details of any of these processes nor could they explain even qualitatively the presence of line emission spectra. Second, the assumption that the electrons revolved about the nucleus was necessary for explaining why the electrons did not follow the force of electric attraction by falling into the nucleus, but this created immediately a seemingly insuperable problem. According to the theory of light emission, any vibrating or revolving charge emits electromagnetic energy, so that each revolving electron should continually spiral closer and closer to the nucleus with ever-increasing speed while emitting light of steadily increasing frequency. As Rutherford said later: "I was perfectly aware when I put forward the theory of the nuclear atom that according to classical theory the electron ought to fall into the nucleus . . ." Of course, both the emission of continuous spectra for ordinary (gaseous) elements and the annihilation of the electrons were predictions in catastrophic conflict with actual fact.

It was at that point that Bohr entered the picture. In essence, he joined the new idea of the nuclear atom with another great recent

concept that had just begun to make its way into scientific thought in Britain, namely, the quantum theory. As Arthur Eddington wrote in 1936: "Let us go back to 1912. At that time quantum theory was a German invention which had scarcely penetrated to England at all. There were rumors that Jeans had gone to a conference on the continent and been converted; Lindemann, I believe, was expert on it; I cannot think of anyone else."

**25-2 Energy levels in hydrogen atoms.** Bohr's original paper was published in the *Phil. Mag.* in July 1913 under the title "On the Constitution of Atoms and Molecules"; other papers followed at short intervals. We cannot here follow Bohr's own line of thought exactly, partly because of its advanced mathematical level, partly because, as even Max Planck himself once exclaimed, Bohr's work is written "in a style that is by no means simple." Nevertheless, the main arguments are quite accessible to us.

Bohr begins by reviewing Rutherford's atomic model and the difficulties it raised concerning electronic orbits and light radiation. Then appears the new theme. Classical electrodynamics, which had originally been confirmed by Hertz's experiments with large circuits creating radio waves, is inadequate for describing the behavior of systems of atomic size. Essentially this is the lesson to be drawn from the work of Planck and Einstein on the quantum theory of light, although they had not analyzed the detailed motion of the atomic "oscillator." The objection to Rutherford's theory that the electron revolving about the nucleus continually loses energy had already been met by Planck, who had proved that the energy in the oscillator does *not* change except discontinuously, in steps of  $hf$ , where  $h$  = Planck's constant and  $f$  = frequency of rotation or vibration.

In this light, consider an atom of hydrogen. It has a nucleus with one positive charge, circled by a single electron at a distance known from kinetic theory to be of the order of  $10^{-8}$  cm. Bohr's first postulate therefore is that while the electron is in a given orbit, which for the time being we may imagine to be circular, *it does not radiate energy*. To radiate a photon, or to absorb energy, the electron must do so while decreasing or increasing its orbit, i.e., during the transition from one *stable* (i.e., nonradiating) orbit to another. Thus to every level on the energy level diagram of the oscillator there corresponds a definite stable orbit, and a sudden transition from one orbit to the next is accompanied by a change of energy level, i.e., from some energy  $E''$  to  $E'$ , so that  $E'' - E' = \Delta E$ ,

which by Einstein's photon theory is equal to the energy of the photon emitted:

$$E'' - E' = \Delta E = h\nu. \quad (25-1)$$

Obviously there are now several questions, above all: What are the actual energies in the different energy levels? Planck's theory only supplied this hint: the energy levels in any one oscillator are all equally spaced, and throughout any sample of radiating atoms there are distributed oscillators of every conceivable type of vibration. These assumptions, as we have seen, led Planck to the correct derivation of the emission curves for glowing solids and liquids, but they had proved powerless for the interpretation of the line emission from gases.

Now Bohr quite boldly denies that for radiating atoms in the gas the same process holds as in solids, liquids, or dense gases. He postulates that in this case *each atomic emitter of hydrogen has identically the same energy level scheme*. Furthermore, these energy levels cannot be evenly spaced; if they were, all hydrogen atoms would send out photons of only one frequency.

We seem to see a way of proceeding. We know the frequencies of light which hydrogen gas emits, therefore we know the energies of the photons. This gives us directly the spacing between energy levels, and so we may construct empirically the energy level diagram. For example, to account for the first four lines in the visible spectrum ( $\lambda = 6563 \times 10^{-8}$  cm,  $4861 \times 10^{-8}$ ,  $4340 \times 10^{-8}$ ,  $4102 \times 10^{-8}$ ), we calculate the energy of these photons ( $h\nu \sim 3.02 \times 10^{-12}$  erg,  $4.07 \times 10^{-12}$ ,  $4.57 \times 10^{-12}$ ,  $4.84 \times 10^{-12}$ ), then build up an

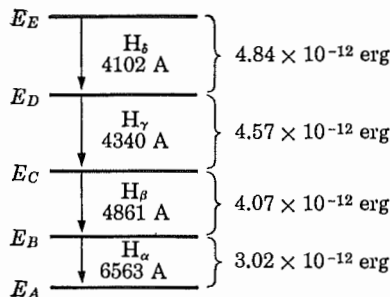


FIG. 25-1. Provisional energy level scheme to account for emission of first four Balmer lines.

energy "ladder" which incorporates these four intervals (Fig. 25-1). Now the emission of the line  $H_\gamma$  ( $\lambda = 4340 \times 10^{-8}$  cm) corresponds to a change of energy level from  $E_D$  to  $E_C$ , and similarly for the others. In principle, we can extend such a ladder to include all observed emission lines.

There occurs a first doubt. Both Planck's and Einstein's work on the quantum theory of light assumed axiomatically that the frequency  $\nu$  of the emitted

light is equal to the frequency  $f$  of oscillation or rotation in the atom. It follows that an electron, while changing from one energy level to the next, must be rotating at a fixed orbital frequency  $f = \nu$ . This is at first sight not inconceivable, for an electron might suddenly spiral from one stable, nonradiating orbit into the next, say to one nearer the nucleus, while retaining its period and its frequency of revolution (Fig. 25-2). But on the one hand this raises the troublesome question of how the electron can thereafter change to a new orbital frequency when it may subsequently have to negotiate the interval from  $E_B$  to  $E_A$  to emit a photon of higher frequency, and on the other hand, one can prove quickly that the electron *cannot* keep a constant frequency or period while it spirals from one stable orbit to the next over any considerable

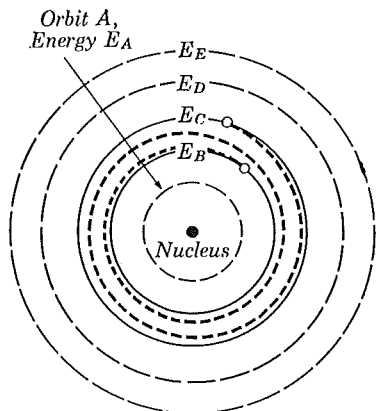


FIG. 25-2. Provisional model of five stable orbits for the electron around the hydrogen nucleus. The electron is spiralling from orbit  $C$  to orbit  $B$ .

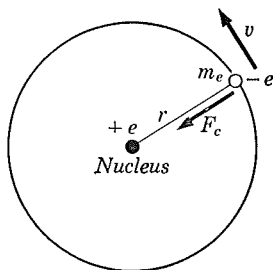


FIG. 25-3.

distance. The net force acting at any instant on the revolving electron, a centripetal Coulomb's law force of electric attraction directed to the central nucleus, as seen in Fig. 25-3, is simply given (numerically) by

$$F_c = \frac{m_e v^2}{r} = \frac{e \cdot e}{r^2}, \quad (25-2)$$

where  $m_e$  = mass of electron,  $e$  = numerical charge on electron and on hydrogen nucleus,  $r$  = distance between electron and nucleus. But the speed of the electron is  $v = 2\pi fr$ , and although this was defined for circular motion, it should hold well enough for motion along portions of a close spiral. Substituting and rearranging terms in Eq. (25-2) yields

$$\frac{1}{f^2} \equiv T^2 = \left( \frac{4\pi^2 m_e}{e^2} \right) r^3. \quad (25-3)$$

In short,  $T^2 \propto r^3$ , nothing less than Kepler's third law! The conclusion is that the period (and frequency) of the electron's orbital revolution does indeed change with the distance from the nucleus, therefore radiation cannot be emitted at constant frequency in the transition between stable orbits. The whole scheme collapses.

There is, however, a daring way of saving the fundamental idea. In the discussion so far, it was assumed that ordinary laws of mechanics and electricity hold during the transition of the electron between energy levels. This led us to a contradiction with experience, therefore Bohr states a second postulate: *Ordinary mechanics and electricity* [e.g., Eqs. (25-2) and (25-3)] *might be used to discuss the motion of the electron in a stable and nonradiating orbit, but "the passing of the systems between different stationary\* states cannot be treated on that basis."* On the contrary, rather than to try any further to visualize what happens during the transition, he asks us to realize that the classical laws were not at all designed to accommodate such intra-atomic processes, that a different mode of thought is required for that region of phenomena, and that the most straightforward approach is to accept a third postulate: during the transition between energy levels, the electron, in a way that cannot be pictured, emits a photon whose frequency  $\nu$  is given by the Planck-Einstein formula ( $\Delta E = E'' - E' = h\nu$ ,  $\therefore \nu = \Delta E/h$ ), *regardless of the frequency  $f$  of orbital revolution of the electron.*

That  $f$  is not equal to  $\nu$  was so completely foreign to all physical thinking that Bohr felt compelled to append this simple remark, "The [last] assumption is in obvious contrast to the ordinary ideas of electrodynamics, but appears to be necessary in order to account for experimental facts." In its impact on the contemporary scientist, it was a statement quite equivalent to Galileo's assertion about three centuries before that all bodies fall at equal rates of descent no matter what their weight. At a later time, Bohr re-emphasized the philosophy of sound scientific pragmatism; classical physics, when applied to the problem of electron stability and line emission, again and again has led to the wrong answers; evidently, the tool is not suited to the problem. "We stand here almost entirely on virgin ground, and upon introducing new assumptions we need only take care not to get into contradiction with experiment. Time will have to show to what extent this can be avoided, but the safest way is, of course, to make as few assumptions as possible."

To summarize, the three radical postulates do give us a scheme

---

\* "Stationary" is Bohr's term for what we have called "stable."

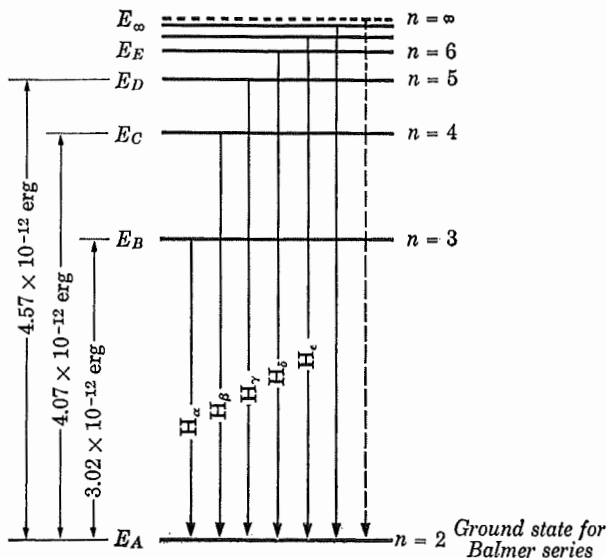


FIG. 25-4. Modified energy level diagram for Balmer series.

for explaining the emission of line spectra by a hydrogen atom (or for that matter, by any other system). We need only to examine all known emission lines for that element, calculate the photon energies, and construct an energy ladder with intervals corresponding to these energies. For each step on that ladder the electron is considered to move in a stable orbit. When the atom absorbs energy from the outside (by the collision with atoms in a hot gas, with particles in electric discharge, with photons, etc.) the electron moves to a new stable orbit with greater energy. On returning to the previous orbit, the energy difference is given up as a photon.

Once it has become clear that we need not account in mechanistic terms for the behavior of the electron during its "jump," we can introduce into the model a profoundly important modification. Instead of spreading out the ladder of energy levels in one long line, with successive transitions allowed only between neighboring levels, we can condense the system as shown in Fig. 25-4 for the lines in the Balmer series. Each jump is made from a higher level to one and the same lowest level, the *ground state* ( $E_A$ ). Here the distance  $E_B - E_A$  represents the energy emitted as a photon of the first (red) Balmer line,  $H_\alpha$ ;  $E_C - E_A$  corresponds to the energy of  $H_\beta$ , and so on. Figure 25-5 is a schematic representation of the jumps negoti-



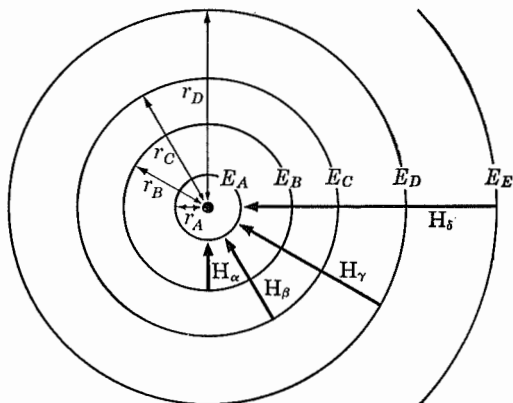


FIG. 25-5. Provisional model for transitions between stable orbits responsible for the emission of the first four Balmer lines.

ated by the electron between its various stable orbits and the ground state.

**PROBLEM 25-1.** Construct an energy level diagram like that in Fig. 25-4 to scale, entering the first six levels and also the level for the highest frequency photon of the Balmer series. (To obtain the necessary data, use the Balmer formula, Eq. (22-3). The highest frequency or shortest wavelength line is at the "convergence limit," i.e., where the integer  $n = \infty$ .) How many energy levels are there between the ground state and the one corresponding to the highest frequency line?

This modification has two immediate consequences. In the previous scheme we required the hydrogen atom to be infinitely dispersed to accommodate the infinitely many stable orbits that it is known to have by virtue of the fact that there are apparently infinitely many lines in the Balmer spectrum. Now, however, we have distributed an infinity of stable orbits in a more rational way, with each jump ending at the same orbit, one that we are tempted to identify with the actual orbit of the electron when the atom is "unexcited," i.e., when it does not have an excess of energy that it may dispose of as a photon. But much more important is the thought that a scheme specifically erected to yield the Balmer spectrum also does more. If an electron in level  $E_C$  does not jump directly to  $E_A$ , but stops at  $E_B$  first, a photon of energy  $E_C - E_B$  (i.e.,  $1.05 \times 10^{-12}$  erg) should be emitted, followed later by one of energy  $E_B - E_A$  (the  $H_\alpha$  line). Similarly, if it came from  $E_\infty$  to  $E_B$  before proceeding to  $E_A$ , a photon of energy  $E_\infty - E_B$  is expected. A

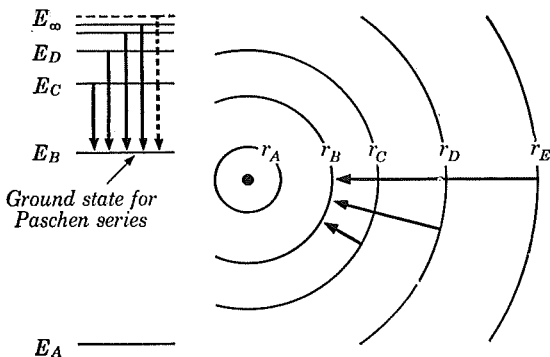


FIG. 25-6. Provisional energy level diagram and orbital transitions for Paschen series.

whole series of lines in the hydrogen spectrum should result by jumps to  $E_B$  instead of  $E_A$ , and each of these lines should differ in frequency from the corresponding Balmer lines by  $(E_B - E_A)/h$ , the frequency of  $H_\alpha$ . This had, in fact, been observed by Paschen in 1908. To fit the Paschen series into the scheme, we do not need to extend the energy ladder by another set of levels, but simply postulate that for the Paschen series the ground state of the electron is  $E_B$  instead of  $E_A$  (Fig. 25-6).

**PROBLEM 25-2.** Calculate the frequencies and wavelengths of the first four (lowest energy) lines in the Paschen spectrum from an energy level diagram made for the Balmer lines of the previous problem.

So far we have filled energy levels empirically and sketched the corresponding orbits qualitatively. Now we shall calculate the radii of the stable orbits. The energy  $E$  of the electron is the sum of its kinetic and (electrical) potential energy.

$$E = \text{KE} + \text{PE} = \frac{1}{2} m_e v^2 + (V_r)(-e). \quad (25-4)$$

This calculation refers to stable orbits, classical laws are presumed to apply, and we see from Eq. (25-2) that  $v^2 = e^2/m_e r$ . Furthermore, the potential  $V_r$  at distance  $r$  from a nuclear charge  $(+)e$  is  $e/r$ , therefore the potential energy is  $-(e^2/r)$ . Rewriting Eq. (25-4),

$$\begin{aligned} E &= \frac{1}{2} m_e \left( \frac{e^2}{m_e r} \right) + \left( \frac{e}{r} \right) (-e) \\ &= \frac{1}{2} \left( \frac{e^2}{r} \right) - \left( \frac{e^2}{r} \right) = -\frac{1}{2} \left( \frac{e^2}{r} \right). \end{aligned} \quad (25-5)$$

The negative sign evidently refers to the fact that work would have to be done *on* the electron to remove it from the positive field of the nucleus to a point at infinity where  $V = 0$ . The sign is simply a consequence of the convention for defining zero potential energy. Since we plan to use Eq. (25-5) only to find *differences* between energy levels, the definition of the zero point will not affect the results.

For the preparation of the energy level diagrams we calculated the quantity  $\Delta E = E'' - E'$ . From Eq. (25-5),

$$\begin{aligned}\Delta E = E'' - E' &= \left[ -\frac{1}{2} \left( \frac{e^2}{r''} \right) \right] - \left[ -\frac{1}{2} \left( \frac{e^2}{r'} \right) \right] \\ &= \frac{1}{2} e^2 \left( \frac{1}{r'} - \frac{1}{r''} \right),\end{aligned}\quad (25-6)$$

where  $r'$  and  $r''$  refer to the radii of the electron orbits with energy  $E'$  and  $E''$ , respectively the states after and before the jump.

**25-3 Correspondence principle.** While the work to this point has been perfectly useful for the next development, we realize that Eq. (25-6) only refers to *differences* in radii, not to the absolute values themselves. It appears that in order to determine the size of the orbits of the electron we must decide on the value of at least one of the radii, essentially by arbitrary postulate. We might argue that the innermost stable orbit should correspond to the dimensions of the unexcited atom, and that the value (of about  $0.5 \times 10^{-8}$  cm) for the effective radius of the atom, known from kinetic theory, should be assigned. (As a problem, you may assume that  $r_A$  in Fig. 25-5 is  $0.5 \times 10^{-8}$  cm, and on that assumption calculate  $r_B$ ,  $r_C$ , and  $r_D$ .) But it is precisely at this point that Bohr's true genius asserted itself most strikingly. He deduced a systematic way of assigning to the hydrogen atom the radii of stable orbits by a consideration of the following type.

If an electron in a strongly excited atom is in a high energy level, its radius is correspondingly large. *But for sufficiently large orbits*, Maxwell's fundamental postulate must hold, by which a circulating current (here only one electron) radiates light energy at a frequency  $\nu$  equal to the frequency  $f$  of the orbital motion. At the same time, Planck's postulate must hold for periodic systems, so that the energy changes of the radiator can proceed only in intervals of  $hf$ . Here is a place where classical and quantum physics join, and we must assume that they overlap without leaving a lawless gap. In this

region the two types of physics merge into each other, and their predictions must correspond. This point of view, or principle, has, however, an astonishing consequence.

In Fig. 25-7, the  $r_1$  and  $r_2$  refer to the orbits of the electron just before and just after it decreased its energy by one quantum,  $hf$ .

The orbits are relatively large by hypothesis, so as to escape the difficulties we previously experienced in connection with Eq. (25-3), By Eq. (25-2) we have

$$r_1 = \frac{m_e v_1^2 r_1^2}{e^2}, \quad r_2 = \frac{m_e v_2^2 r_2^2}{e^2},$$

and Eq. (25-6) becomes

$$\Delta E = \frac{1}{2} e^4 \left[ \frac{(m_e v_1 r_1 + m_e v_2 r_2) \cdot (m_e v_1 r_1 - m_e v_2 r_2)}{m_e v_2^2 r_2^2 \cdot m_e v_1^2 r_1^2} \right]. \quad (25-7)$$

Since  $r_1 - r_2 (\equiv \Delta r)$  is very small,  $r_1 \doteq r_2 (\equiv r)$ ,  $v_1 \doteq v_2 (\equiv v)$ . This, together with  $v^4 = e^4/m_e^2 r^2$  from Eq. (25-2), reduces the expression (show this!) to  $\Delta E = v \Delta(m_e v r)/r$ . Because  $v = 2\pi n r$ , where  $n$  is the frequency of orbital rotation of the electron,  $\Delta E = 2\pi n \Delta(m_e v r)$ . Use of Eq. (25-1) leads to  $2\pi n \Delta m_e v r = h\nu$ . Now we invoke the correspondence principle: for sufficiently large orbits, the atomic system we are dealing with will obey classical laws. If so, the frequency  $\nu$  of emitted radiation should be numerically equal to the frequency  $n$  of the orbiting electric charge. The last expression then becomes

$$\Delta(m_e v r) = \frac{h}{2\pi}. \quad (25-8)$$

This is extremely important: The *change* in the quantity  $(m_e v r)$  of the orbiting electron as a result of a postulated very small decrease in its orbital radius is discovered to be a *universal constant*. Moreover, note that  $m_e v r$  is simply the angular momentum  $I\omega$  of the electron. This invites the speculation that the whole angular momen-

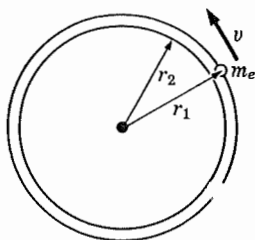


FIG. 25-7.

tum of the electron, anywhere, in any of its possible orbits, can be thought of as some integral multiple  $n$  of a smallest permitted value given by Eq. (25-8), i.e., that  $m_e v r = n \cdot \Delta m_e v r$ ; and by Eq. (25-8),

$$m_e v r = n \frac{h}{2\pi}, \quad \text{or} \quad r = \frac{n h}{2\pi m_e v}, \quad (25-9)$$

where the letter  $n$  now stands, of course, not for frequency, but for an integer (1, 2, 3, . . .).

From Eq. (25-2), which describes the condition that must exist in any stable orbit, we have

$$r = \frac{e^2}{m_e v^2}. \quad (25-10)$$

But Eq. (25-9) gives for the square of the velocity,

$$v^2 = \frac{n^2 h^2}{4\pi^2 m_e^2 r^2}. \quad (25-11)$$

Substituting the expression for  $v^2$  in Eq. (25-10) gives

$$r = \frac{n^2 h^2}{4\pi^2 m_e e^2}, \quad (25-12)$$

where  $h$ ,  $m_e$ , and  $e$  are the usual constants, and  $n = 1, 2, 3$ , etc. In Eq. (25-12) we have a very definite directive. If the angular momentum is "quantized," i.e., subdivisible in whole multiples of  $(h/2\pi)$  only, then the orbital radii are quantized also. That this had to be so we know well, for unless the orbits are restricted to definite radii we do not expect to obtain definite and distinct energy levels. The remarkable point is that Eq. (25-12) prescribes a set of radii without referring at all to experimental observations of line spectra. In fact, we can now obtain numerical values for the "permitted" radii. When  $n = 1$ ,  $r$  is smallest, and has in fact the value

$$r = 0.53 \times 10^{-8} \text{ cm.}$$

This is just the order of magnitude we expect for the dimensions of the unexcited atom, i.e., the atom in its *natural state*. If  $n$  has a higher value, the electron must be thought to be in a higher energy level, and the orbital radius is proportional to  $n^2$ . Because the values of the integer  $n$  determine the quantization of radii  $r$ , we can call  $n$  the *quantum number*. Figure 25-8 presents schematically the first few stable orbits corresponding to the first few quantum numbers.

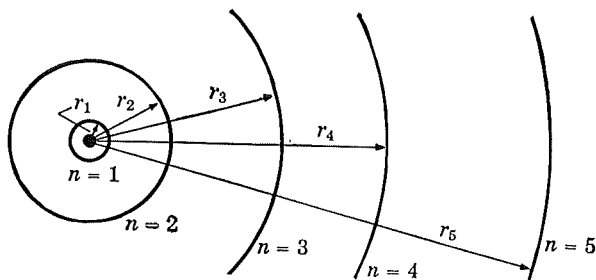


FIG. 25-8. First few stable Bohr orbits.

**PROBLEM 25-3.** Construct to scale the first five “permitted” or stable circular orbits ( $n = 1$  through 5) for the hydrogen atom.

**25-4 The explanation of hydrogen line emission spectra.** So far, the idea of quantization of orbits has given us a set of radii that is plausible at least as far as the innermost radius ( $n = 1$ ) is concerned. Now we must test whether we can put this information to the use initially intended, namely, to the prediction of energy level differences in accord with experience. Equation (25-5) has given us the proper expression for each energy level, which we may now combine with Eq. (25-12):

$$E = -\frac{e^2}{2r} = -\frac{e^2}{2\left(\frac{n^2\hbar^2}{4\pi^2m_e e^2}\right)} = -\frac{2\pi^2m_e e^4}{n^2\hbar^2}. \quad (25-13)$$

This at last enables us to calculate the energies in each level, and to construct an energy level diagram.

**PROBLEM 25-4.** Calculate and construct an energy level diagram, starting with the level for which the quantum number  $n = \infty$ , then for  $n = 1, 2, 3$ , and 4. What is the meaning of the negative values?

The real interest centers on the differences between energy levels, for this is what corresponds to observations when there is photon emission. If the steps leading to Eq. (25-13) are correct, the frequency of the photon emitted when the electron jumps from a distant orbit ( $n''$ ) to an orbit nearer to the nucleus ( $n'$ ), should be given by

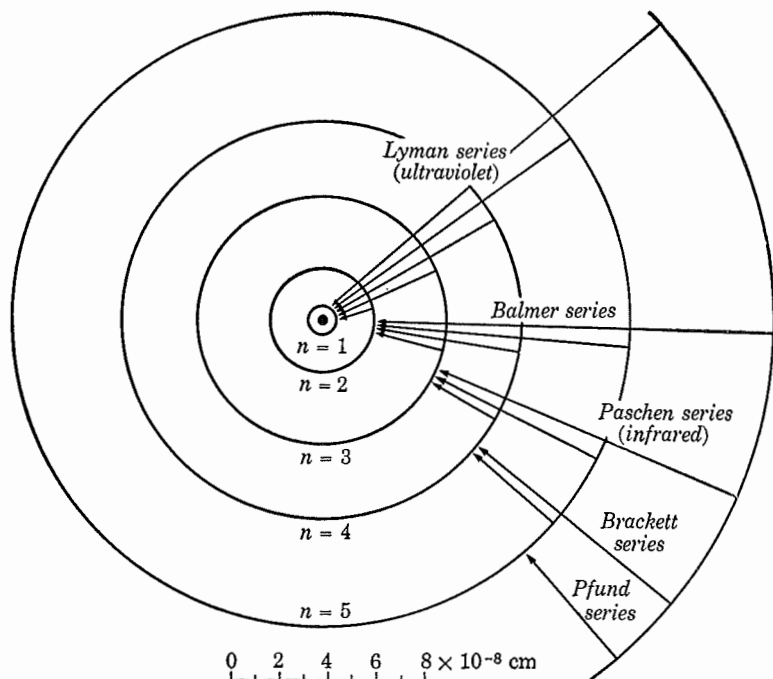


FIG. 25-9. Final scheme of transitions between stable Bohr orbits in hydrogen atom (approximately to scale).

$$\begin{aligned} \nu &= \frac{E'' - E'}{h} = \left[ -\frac{2\pi^2 m_e e^4}{(n'')^2 h^3} \right] - \left[ -\frac{2\pi^2 m_e e^4}{(n')^2 h^3} \right] \\ &= \frac{2\pi^2 m_e e^4}{h^3} \left[ \frac{1}{(n'')^2} - \frac{1}{(n')^2} \right]. \end{aligned}$$

In terms of the more directly measurable wavelength  $\lambda$ , this becomes

$$\frac{1}{\lambda} = \frac{2\pi^2 m_e e^4}{h^3 c} \left[ \frac{1}{(n'')^2} - \frac{1}{(n')^2} \right]. \quad (25-14)$$

We recognize at once that this *derived* expression is indeed identical with the *empirical* formula for the hydrogen spectra [Eq. (22-3)]

$$\frac{1}{\lambda} = R \left[ \frac{1}{(n'')^2} - \frac{1}{(n')^2} \right], \quad (25-15)$$

provided that (a) the combination of constants  $2\pi^2 m_e e^4 / h^3 c$  is equal to  $109,678 \text{ cm}^{-1}$ , Rydberg's empirical constant  $R$  (which indeed is

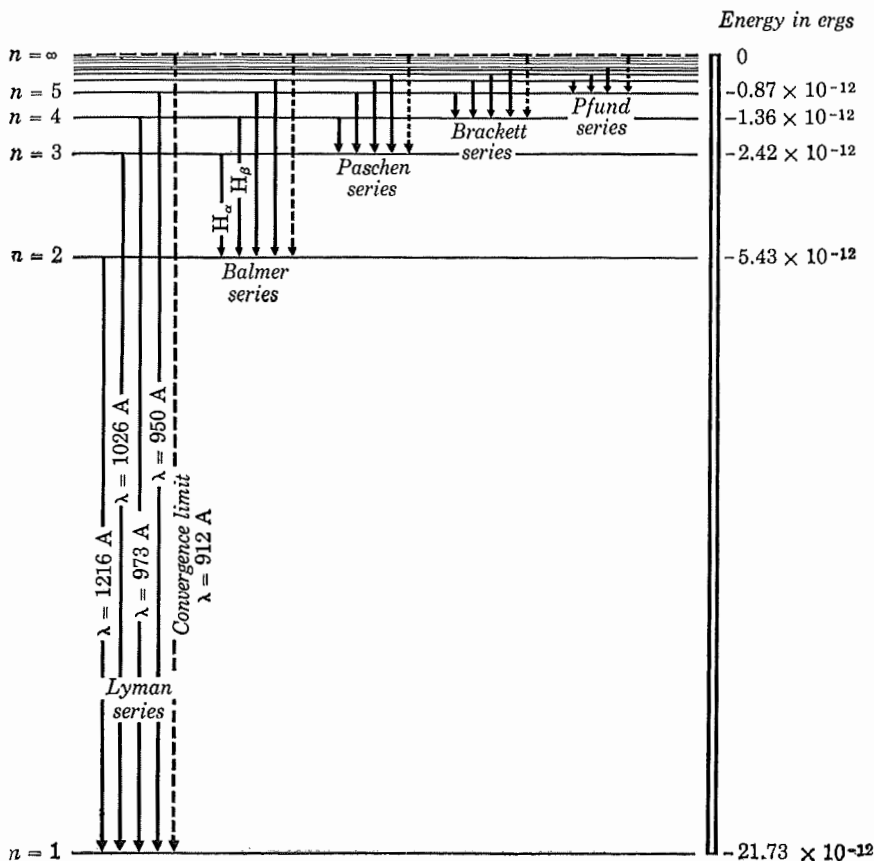


FIG. 25-10. Energy level diagram for hydrogen (final form).

the case within 0.06%\*), and (b) that we may interpret the quantum numbers  $n'$  and  $n''$  according to the following scheme. When  $n' = 2$  and  $n'' = 3, 4, 5, \dots$ , which gives us the lines of the Balmer series from the empirical Eq. (25-15), the ground state for all transitions is the second smallest orbit (see Fig. 25-9). When  $n' = 3$  and  $n'' = 4, 5, 6, \dots$  (the condition that is met for the Paschen series), the ground state is the third orbit, and so on for the other series.

For example,  $n' = 1, n'' = 2, 3, 4, \dots$  refers to the Lyman series, which was not fully identified until 1914, but which in 1913 was generally expected to exist in the ultraviolet. Figures 25-9 and

\* Using modern values. In early 1913, the discrepancy was 6%, yet the agreement was very impressive.



25-10 show where in the orbital scheme and on the energy level diagram the Lyman series fits in with the others. The radii and energy levels for this series can be calculated, independent of spectroscopic data, from Eqs. (25-12) and (25-13) by assigning  $n = 1$  (the ground state is here the natural state). Note well that nothing of the kind was possible in our initial, unsuccessful attempt to construct energy level diagrams and orbital schemes from the available experimental data of line emission. Lacking lines from the Lyman series, we could not at that time avoid the (incorrect) assumption that the ground state for transitions connected with Balmer series emission was the lowest natural state; only with an independent postulate, such as that of quantization of angular momenta, can one hope to fix the radii before all the lines are known.

Thus we have come to the initial successes of Bohr's model. It predicts the size of the unexcited atom, and the prediction is of the correct order of magnitude. It yields an expression for the wavelengths of all lines that are to be expected from radiating hydrogen atoms, and this expression coincides exactly with one that summarizes the experimental facts of line emission. It accounts for Rydberg's empirical constant in terms of physical quantities. It provides us with a visualizable (although therefore perhaps dangerous) system, and establishes a physical order among the events accompanied by emission, whether the particular lines are well known or as yet beyond the region of experimentation. The model introduces quantum theory into the atom and thereby gives on the one hand a physical basis for Planck's induction that the energy in the atom is quantized and on the other hand removes the problem of the stability of electron orbits from the classical theory which could not provide the solution. And as hinted in Section 25-3, it emphasizes through the correspondence principle a point of view that avoids the complete separation of classical and quantum physics into two unrelated fields.

PROBLEM 25-5. With proper techniques, each of these four lines may be found in the hydrogen spectrum:  $\lambda = 1215.7 \text{ \AA}$ ,  $\lambda = 3970.1 \text{ \AA}$ ,  $\lambda = 9500 \text{ \AA}$ ,  $\lambda = 74,000 \text{ \AA}$ . Attempt to identify in each case to which series the line belongs, and fix the number of the orbits (or the value of the quantum numbers) involved in the particular transition.

Now that we have let the development of Bohr's model take us by the hand, as it were, and lead us to the postulates and through the derivations, we may briefly review how one can obtain the same

final result more quickly if one no longer needs to convince oneself of the necessity and full import of individual steps. In fact, the following sequence has frequently been given, even by Bohr in his subsequent publications:

(a) Begin by assuming that the angular momentum of the electron orbit is quantized according to  $I\omega = nh/2\pi$ . (b) Therefore the radii and energies are quantized according to Eqs. (25-12) and (25-13). (c) Photons of frequency  $\nu$  and of energy  $h\nu = E'' - E'$  are emitted in the discontinuous transition from any one orbit with energy  $E''$  to any other with energy  $E'$ . (d) From the last assumption follows Eq. (25-14), the prediction of wavelengths in line emission from hydrogen. (e) Comparison with the experimental facts, summarized in Eq. (25-15), proves the accuracy of the last conclusion, and thereby establishes the model that led to it.

**PROBLEM 25-6.** Provide a derivation of Eq. (25-14) by following the steps just indicated, but point out specifically what postulates or assumptions that deviate from the classical theory are used.

**PROBLEM 25-7.** Verify by calculation the agreement between Rydberg's constant and the expression in Eq. (25-14). (For data refer to Appendix I.)

**25-5 Further results in Bohr's original paper.** While the derivation of the empirical hydrogen spectrum formula was a superb feat in itself, Bohr also extended his argument from the very beginning to encompass a large variety of problems—rather like Newton, who followed his explanation of Kepler's empirical laws with a discussion of the inequalities in the moon's motions, the problem of comets, of the tides, and the like. In fact, it is the hallmark of great conceptual schemes that their applicability far exceeds the boundaries of initial intent, as has been amply illustrated in the preceding cases. The following additional problems were discussed in Bohr's original paper:

(a) *Ionization potential.* From Bohr's model it followed that the energy  $E$  of the hydrogen atom's electron in its innermost stable orbit ("natural state") is given by Eq. (25-13), where  $n = 1$ , i.e.,  $E = -2.18 \times 10^{-11}$  erg. To put it another way, the electron must be given  $2.18 \times 10^{-11}$  erg of energy to allow it to move (against the force of attraction to the nucleus) from the first orbit to infinity, or to a large enough distance from the nucleus so that the remaining energy is virtually zero. But this state of affairs, the loss of the electron, corresponds to the ionization of the hydrogen atom, and therefore Bohr's theory predicts that one must supply  $2.18 \times 10^{-11}$  erg to

a hydrogen atom to cause its ionization. However, the actual energy requirement for ionizing atoms can be measured experimentally by accelerating a beam of electrons into a tube filled with a little hydrogen gas. If the electrons in the beam are sufficiently energetic, during a collision they will impart to the orbital electrons in hydrogen atoms enough energy to allow the orbital electrons to leave their parent atoms entirely, which in turn will cause the gas to become suddenly ionized and highly conducting. We find it takes an accelerating voltage of 13.6 volts (or 13.6/300 statvolts) for the electron beam to induce a sudden change in conduction through, or ionization in the gas tube. But with that accelerating voltage, each projectile, having the electronic charge  $e$ , carries an energy of  $(13.6/300) \times (4.8 \times 10^{-10})$  or  $2.17 \times 10^{-11}$  erg, in excellent agreement with the prediction from Bohr's theory.

**PROBLEM 25-8.** When a completely ionized atom of hydrogen captures an electron, the latter may conceivably cascade toward the nucleus by jumping successively from one permitted orbit to the next. (a) If we could watch the light emitted by this single atom, to what lines of the complete hydrogen spectrum will the last four jumps correspond? (b) Why will the electron not jump from the innermost orbit ( $n = 1$ ) to the nucleus? (Be cautious in answering this last question.)

(b) *Number of emission lines.* Balmer's formula and Bohr's derivation seem to agree also that infinitely many lines in the Balmer series should be observable in principle. Yet it had long been known that in gas discharge tubes in the laboratory only the first 10 or 12 lines are obtainable, compared with 33 in the spectra of celestial bodies (sun's corona). This "is just what we should expect from the above theory." The higher the frequency of the emitted line, the larger the jump of the orbital electron; to generate the 33rd Balmer line, the electron must move in toward the second stable orbit from a radius of  $6 \times 10^{-6}$  cm [see Eq. (25-12),  $n = 34$ ]. But in ordinary discharge tubes, an electron will generally not be able to move out to such a large orbit without other atoms interfering. Even at the relatively low pressures in the tube the density of hydrogen is about  $5 \times 10^{14}$  atoms per  $\text{cm}^3$ , and consequently the average distance between atoms is about  $10^{-5}$  cm, of the order of the atomic diameter corresponding to  $n = 34$ . "According to the theory, the necessary condition for the appearance of a great number of lines is therefore a very small density of the gas; for simultaneously to obtain an intensity sufficient for observation, the space filled with

the gas must be very great." This double condition of low density and large volume is fulfilled only in the vast regions of stellar atmospheres.

PROBLEM 25-9. Just how many hydrogen atoms are there in a 5 cm<sup>3</sup> discharge tube if it was filled with H<sub>2</sub> gas at 0°C and pumped down to a pressure of  $2 \times 10^{-5}$  atm before the current is passed through? What is the average distance between atoms?

(c) *Spectrum of ionized helium.* Bohr also points out that his model and his derivation do not provide for the lines, then generally ascribed to hydrogen, which E. C. Pickering had described in 1896-97 in stellar spectra. In 1912, A. Fowler also found these lines in discharges through mixtures of hydrogen and helium. Their wavelength is given by

$$\frac{1}{\lambda} = R \left( \frac{1}{(n'/2)^2} - \frac{1}{(n''/2)^2} \right). \quad (22-5)$$

However, Bohr continues, "we can account naturally for these lines if we ascribe them to helium [specifically, singly ionized helium, He<sup>+</sup>]. A neutral atom of the latter element consists, according to Rutherford's theory, of a positive nucleus of charge  $2e$  and two electrons." Now note that if one of the electrons is stripped off through the violence either of the electric discharge or of mutual atomic collisions in hot stars, the remaining helium ion is just like a hydrogen atom *except* for a doubly positive central charge. The light emitted by transitions of the single remaining electron can be calculated at once.

The centripetal force  $F_c$  on the electron is now given numerically by

$$F_c \equiv \frac{m_e v^2}{r} = \frac{e(2e)}{r^2},$$

i.e., by Coulomb's law. Therefore the condition fulfilled by the remaining electron in every stable orbit is

$$r = \frac{2e^2}{m_e v^2} \left( \text{or } v^2 = \frac{2e^2}{m_e r} \right).$$

The angular momentum of an electron is always assumed to be quantized:

$$m_e v r = \frac{n\hbar}{2\pi}, \quad \text{or } v^2 = \frac{n^2 \hbar^2}{4\pi^2 m_e^2 r^2}.$$

Combining the last expression for  $v^2$  and  $r$ , we have for the permitted radii

$$r = \frac{n^2 h^2}{4\pi^2 m_e (2e^2)}, \quad (25-16)$$

which differs from Eq. (25-12) for hydrogen atoms by the factor 2 in the denominator. Now we calculate the energies in the allowed orbits of  $\text{He}^+$ :

$$\begin{aligned} E &= \text{KE} + \text{PE} = \frac{1}{2} m_e v^2 + V(-e) \\ &= \frac{1}{2} m_e \left( \frac{2e^2}{m_e r} \right) + \left( \frac{2e}{r} \right) (-e) \\ &= -\frac{e^2}{r}. \end{aligned} \quad (25-17)$$

From Eqs. (25-16) and (25-17), we derive at once

$$\frac{1}{\lambda} \equiv \frac{\nu}{c} = \frac{E'' - E'}{hc} = \frac{2\pi^2 m_e (2)^2 e^4}{h^3 c} \left[ \frac{1}{(n')^2} - \frac{1}{(n'')^2} \right], \quad (25-18)$$

which we may write

$$\frac{1}{\lambda} = \frac{2\pi^2 m_e e^4}{h^3 c} \left[ \frac{1}{(n'/2)^2} - \frac{1}{(n''/2)^2} \right]. \quad (25-19)$$

But the last equation coincides with the observations expressed in Eq. (22-5), since it has been proved that Rydberg's constant  $R$  is equivalent within small limits to the factor outside the bracket! Bohr says: "If we put  $n' = 3$ , and let  $n''$  vary [i.e.,  $n'' = 4, 5, \dots$ ], we get a series which includes 2 of the [lines] observed by Fowler . . . If we put  $n' = 4$ , we get the series observed by Pickering in the spectrum of  $\zeta$  Puppis. Every second of the lines in this series is identical with a line in the Balmer series of the hydrogen spectrum." This is indeed an exciting interpretation of an old puzzle.

It is clear that Eq. (25-18) differs from Eq. (25-14) only by the factor  $(2)^2$ . If an atom with  $Z$  positive nuclear charges were stripped of all but one electron, Eq. (25-18) should apply to its spectrum except that  $(2)^2$  is replaced by  $Z^2$ . And indeed, it is possible to produce such violent discharges that not only helium but each of the light elements up to the 8th, oxygen, is suitably ionized. The lines emitted from such gases are in full agreement with Bohr's theory.

**PROBLEM 25-10.** The vapor of lithium ( $Z = 3$ ) may be doubly ionized ( $\text{Li}^{++}$ ). Derive the equation analogous to Eq. (25-19) for the line spectrum

from  $\text{Li}^{++}$ , and draw an energy level diagram and a diagram of the first few orbits analogous to Figs. 25-9 and 25-10.

(d) *Absorption of radiation.* We must now see whether Bohr's model accounts for absorption of light as well as it did for emission. Two facts have been noted in the discussion of line absorption spectra: (i) that light with a continuous spectrum, on passing through a gas, may lose frequencies that correspond exactly to those the gas would itself emit in a line emission spectrum, and (ii) that the absorption spectrum of a gas is, however, usually less rich than its emission spectrum. In this specific case, monatomic hydrogen gas generally absorbs only lines corresponding to the Lyman spectrum, as we now know; it has to be brought to an excited state if it is to absorb the lines corresponding to the Balmer spectrum, as Bohr pointed out. If we make the assumption, so consistent with the whole model, that an orbital electron can absorb a photon only if the energy so absorbed brings it out exactly into another permitted orbit, the whole set of observations becomes clear. Ordinarily, hydrogen atoms are in the most stable, the "natural," state corresponding to  $n = 1$ . The energy differences between that level and all others correspond to lines in the Lyman spectrum; photons with less energy, as all those in the Balmer spectrum, cannot be absorbed because they are unable to make the electron go even from the first innermost orbit to the second one. Only if the gas is initially in an excited state will sufficient atoms be present in the particular ground state of the Balmer series ( $n = 2$ ) to absorb Balmer series frequencies.

On the same model, another puzzle may be solved. It had been noted that above a certain frequency of incident light there is no longer *line* absorption, but *continuous* absorption of all frequencies. Evidently a photon that corresponds to the highest frequency line in any series is one that, if absorbed, produces complete ionization of the hydrogen atom by carrying the electron to the level  $n = \infty$ . That electron is then no longer bound to the atom but is free; but for free electrons there is no energy quantization, they can accept and carry (as translational kinetic energy) any amount of energy whatever. Evidently, when a photon with energy larger than sufficient to free the electron is absorbed, part of the energy is used to ionize the atom, the rest is given to the electron as kinetic energy.

We realize now that we have actually been speaking about the photoelectric effect as applied to gases: only photons above a threshold frequency  $\nu_0$  will cause photoelectric emission, and  $\nu_0$  is  $\Delta E/h$ , where  $\Delta E$  is the difference of energy levels between the ground

state of the illuminated gas and the state for which  $n = \infty$ . Bohr says "Obviously, we get in this way the same expression for the kinetic energy of an electron ejected from an atom by photoelectric effect as that deduced by Einstein, i.e.,  $\text{KE} = h\nu - \mathcal{W} \dots$ "

PROBLEM 25-11. A sample of monatomic hydrogen gas is illuminated from a source emitting only light of wavelength  $\lambda = 1.026 \times 10^{-6}$  cm (second Lyman line). If Bohr's picture of absorption and radiation is correct, and if the hydrogen atoms later re-emit the energy that they absorb, what *two* wavelengths of light do we expect to be re-radiated from the sample?

PROBLEM 25-12. Calculate the work function  $\mathcal{W}$  for the photoelectric effect in unexcited monatomic hydrogen gas.

PROBLEM 25-13. As Bohr was aware, Rutherford had noted in 1912 that fast  $\beta$ -rays sent through a gas "will lose energy in distinct finite quanta." Explain this observation in terms of Bohr's model.

PROBLEM 25-14. According to the kinetic-molecular theory of gases, the average kinetic energy per gas molecule is given by  $\overline{\text{KE}}_{\text{trans}} = \frac{3}{2} kT$ , where  $k =$  Boltzmann's constant,  $1.380 \times 10^{-16}$  erg-deg $^{-1}$ , and  $T =$  temperature in  $^{\circ}\text{K}$  [see Eq. (20-11)]. If the orbital electron can be thrown into a higher energy level by absorbing the kinetic energy during an inelastic collision between two atoms, calculate the temperature of the gas (for monatomic hydrogen) at which most gas atoms are in an energy level where  $n = 2$  and consequently can contribute to the prominent absorption of lines with the Balmer series frequencies. (*Note:* even at a lower temperature, some fraction of the total number of atoms will be sufficiently excited to aid in this type of absorption. Why?)

(e) *Multi-electron systems.* To this point, Bohr's concern has been only with single-electron systems, e.g., hydrogen and ionized helium. Such physical properties of hydrogen atoms as dimensions, emission spectra, absorption spectra, and ionization potential having been explained with brilliant success, the next problem is evidently that of atoms whose nuclei are surrounded by more than one electron. Here the initial problem is again twofold: first, one must determine some method for obtaining electron configurations that are stable, then one must show that the motion of these electrons accounts quantitatively for such observations as the line emission spectra. As to the first problem, Thomson's work on electronic ring structures during the previous years seemed to have given an answer, but it was incapable of providing for spectra, being based on purely classical physics and conceived in terms of very mechanical analogies such as the configurations assumed by a group of magnets fastened

to corks on water.\* Moreover, Bohr now points out that Thomson's model is after all not stable even on the basis of classical theory. Instead, Bohr proposes to carry over from the analysis of the single-electron system the hypothesis that the angular momentum of the electrons is quantized, and add to it the postulate that the configuration of electrons around the nucleus is such that the energy of the system is smallest. In Bohr's words "we shall assume that a configuration . . . is stable if the total energy of the system is less than in any neighboring [i.e., slightly differently arranged] configuration satisfying the same condition of the angular momentum of the electrons." In practice, these conditions pose great difficulties in calculation, because in determining the energy of a system of several bodies one must include the effect on each by all the others. But this was precisely the main problem Bohr put himself in Part II of his main paper, published two months after Part I.

The picture developed there (one whose subsequent modification we shall follow) is that of a series of coplanar concentric rings of electrons. Bohr considers how a nucleus would allow electrons to arrange themselves about it if they were supplied to the system separately, one after the other. The first result, one of considerable importance, is that the neutral hydrogen atom may, under certain conditions, hold another electron in a stable configuration in the innermost ring around the nucleus; thus a neutral hydrogen atom can fairly readily accommodate an extra electron, thereby becoming negatively ionized ( $H^-$ ) by snatching an easily detachable electron from another atom. However, Bohr's calculations show that it "cannot be expected to be able to acquire a double negative charge [ $H^{--}$ ]." A doubly positive nucleus, e.g., a helium nucleus, would force its two orbital electrons to move at equal radii about one-half that of the hydrogen electron. The calculated ionization potential for He is very high, 27 volts, indicating the stability of the system. It

\* It is significant that the same issue of the *Phil. Mag.* in which Bohr's paper appeared carried also a long and very thorough elaboration of Thomson's idea by one of his followers, a discussion of the structure of atoms and molecules based partly on experiments made with small charged spheres. No hint of Planck's or Einstein's work appears. In another paper in the same volume, a prominent British mathematical physicist ends his paper with the plea: "The quantum introduces into physics ideas which seem incapable of coming to any kind of terms with ordinary mechanical concepts. I suggest that we ought rather to seek some system in which as much as possible of the old may be retained." On reading through this one volume of the *Phil. Mag.*, one can virtually feel the excitement and the agony of a science in revolution.



is in good agreement with the observed value. A third electron would generally not find a stable position around the neutral helium atom, "this is in agreement with experimental evidence, which shows that helium atoms have not 'affinity' for free electrons."

You will note that Bohr's work is beginning to explain the chemical behavior of elements! Hydrogen may gain or lose one electron (this we should interpret to mean its valence is 1), but helium has two and only two electrons in a stable system (a valence of zero, an "inert" gas). While these conclusions were not elaborated upon at the time, they eventually proved to be as impressive a consequence of Bohr's models as any discussed so far.

After helium, the next atom discussed is lithium. Several different configurations of three electrons around the nucleus would fulfill the conditions of the orbital angular momentum, and Bohr proposes one that allows two electrons in one orbit quite near the nucleus, and the third in a large orbit around the central system. This outer electron is bound to the nucleus "even more lightly than the electron in a hydrogen atom," and can be detached by relatively small expenditures of energy. "This may have a bearing on the explanation of the apparent tendency of lithium atoms to take a positive charge in chemical combinations with other elements."

In the sequel, Bohr shows that the electrons for elements of higher atomic number will separate into several concentric rings. The complexity of calculations being as great as the choice among equivalent arrangements, he suggests that the number of electrons in the outermost ring be "arbitrarily put equal to the normal valency of the corresponding element." Thus sodium, with valence 1, is thought to have two concentric rings, one with two electrons, the other with eight, and the eleventh electron in a third ring at some distance from the nucleus. This fruitful thought is at once developed further: "At the end of the third period of 8 elements we meet with the iron-group [Fe, Co, Ni]. This group takes a particular position in the system of the elements, since it is the first time that elements of neighboring atomic weights show similar chemical properties. This circumstance indicates that the configurations of the electrons in the elements of this group differ only in the arrangement of the inner electrons . . . On the same lines, the presence of the group of the rare earths [which are chemically similar] indicates that for still greater values of  $Z$  [central nuclear charge] another gradual alteration of the innermost rings will take place." These superb intuitions were subsequently also borne out.

(f) *Characteristic x-rays.* Bohr next turns to a suggestion, made previously by Thomson, that whereas the visible line spectra of elements are emitted by the orbital transitions of external electrons, the characteristic x-ray spectrum ( $K$ ,  $L$  lines) "is sent out during the settling down of the system if electrons in the inner rings are removed by some agency, e.g., by impact of cathode particles," that is, by an electron beam. He makes a very rough calculation of the speed that an electron in a cathode ray must have to remove an inner electron from its site near a nucleus of charge  $Ze$ . The centripetal force holding it, neglecting the effect of all other electrons, is by Coulomb's law

$$F_c = \frac{(Ze)e}{r^2} = \frac{m_e v^2}{r}.$$

But by the rule of quantization of angular momentum for the *innermost* orbit (where  $n = 1$ ),

$$m_e v r = \frac{h}{2\pi}.$$

Substituting, Bohr obtained for the speed of the innermost electron

$$v = \frac{2\pi Z e^2}{h} = 2.1 \times 10^8 Z \text{ cm/sec.} \quad (25-20)$$

However, the total energy of a bound orbital electron is numerically equal to its KE [see Eq. (25-5)]. Therefore, to remove it from the atom, that much energy, or more, must be supplied; since the energy here is supplied by collision with an object of equal mass, the latter must approach at least at an equal speed. And indeed, it had been found experimentally in 1911 that just such an equation as (25-20) gave the correct minimum speed of cathode rays for the generation of the short-wavelength  $K$  lines of x-rays from targets of different elements.

This new confirmation was very gratifying. Moreover, within another three months, Moseley's paper appeared. This we have already discussed briefly. There the argument was carried on in a beautiful way. Moseley assumed that the  $K_\alpha$  line is emitted when, subsequent to the violent removal of an electron from the innermost ring, one from the next ring falls into its place. (This vacancy is then filled in turn by an electron from the following ring, with the emission of a longer-wavelength line, the  $L$  line, and so forth.) As far as the first jump is concerned, Moseley considers that the electron is faced with a nucleus of  $Ze$  charges, partially screened by

whatever electrons still remain in the innermost ring. There is good reason to assume an effective central charge of  $(Z - 1)e$ , particularly now that it is clear to us that the innermost ring has one orbital electron remaining after the cathode-ray collision. At once one is tempted to use for the wavelength of the photon emitted in the transition the same derivation and the same formula as Eq. (25-18), except that now the effective central charge, namely (2), is replaced by  $(Z - 1)$ ; here  $n' = 1$ ,  $n'' = 2$ , this being a transition between the two innermost possible orbits. This gives us the equation for the wavelength of  $K_\alpha$ :

$$\frac{1}{\lambda} = \frac{2\pi^2 m_e e^4 (Z - 1)^2}{h^3 c} \left( \frac{1}{1^2} - \frac{1}{2^2} \right). \quad (25-21)$$

Using as an example  $K_\alpha$  for copper ( $Z = 29$ ), we obtain  $\lambda = 1.55 \times 10^{-8}$  cm, in remarkable agreement with  $\lambda \doteq 1.54 \times 10^{-8}$  cm (see Table 24-4).

We may rephrase Eq. (25-21) to express frequencies by substituting modern values:

$$\nu_{K_\alpha} = 2.47 \times 10^{16} (Z - 1)^2, \quad (25-22)$$

again in agreement with the *empirical* Eq. (24-3). Note once more that the  $K_\alpha$  lines represent a property of atoms without periodicity, i.e., that  $\nu$  is proportional to  $(Z - 1)^2$  in regular ascent through the periodic table.

Moseley noted that the Bohr theory was indeed most successful in explaining the  $K_\alpha$  line, but that "no explanation of the faint line [ $K_\beta$ ] has been forthcoming. Probably further experiments will show that the theory needs some modification," as indeed was the case.

**PROBLEM 25-15.** Extend the previous argument to find the wavelength of the  $L$  line of copper emitted when an electron from the third ring jumps to the vacancy in the second ring. Assume that there are two electrons in the first, and eight in the second ring when each is filled. ( $\lambda$  of  $L_\alpha$  is actually  $13.3 \times 10^{-8}$  cm.)

(g) *Origin of beta rays.* While it was clear at the time of Bohr's writing that the  $\alpha$ -particles emitted during radioactive decay came from the nucleus, the origin of  $\beta$ -radiation was not settled. To many, it seemed logical to assume that a  $\beta$ -ray is essentially an electron ejected at high speed from the cloud of *extranuclear* electrons of the heavier, radioactive atoms. But Bohr now pointed out that his atomic model, substantiated at least in outline by the evidences

at hand, required that the arrangement of the electron cloud be in orbits of greatest stability and least energy. There is no way in which an orbital electron can suddenly, of itself, attain the necessary energy to leave the atom at the enormous speeds that  $\beta$ -rays exhibit. Furthermore, the idea of isotopes had just become plausible, and whereas the identity of chemical behavior of different isotopes of the same element would seem to demand identical structure of the extranuclear electron cloud, the velocity of  $\beta$ -rays from two isotopes could be quite different (e.g.,  $\text{Th}^{234}$ ,  $\beta = 0.3$  Mev;  $\text{Th}^{231}$ ,  $\beta = 0.2$  Mev). Both arguments pointed clearly to a nuclear mechanism of  $\beta$ -ray emission.

This material concluded Part II of Bohr's paper. Part III appeared in November 1913, and dealt with an extension of the model to the explanation of how atoms associate to form molecules by the distribution of electron clouds about the two nuclei. However, we have had sufficient evidences to illustrate the magnitude and inherent fruitfulness of Bohr's initial contribution. Although one outstanding problem was so far untouched (that of the spectra of multi-electron systems), and although soon the model of the atom had to undergo further fundamental modifications, partly at the hand of Bohr himself, nevertheless the picture given so far is the basic scheme on which our present understanding of the atom rests. The praise justly heaped upon the shy young man when the importance of his work began to be realized is reflected in Rutherford's comment:

"Bohr's original quantum theory of spectra was one of the most revolutionary, I suppose, that was ever given to science, and I do not know of any theory that has been more successful . . . I consider the work of Bohr one of the greatest triumphs of the human mind."

PROBLEM 25-16. List the available contemporary experimental facts and conceptual schemes useful to Bohr at the time of the formulation of his atomic model.

**25-6 Further developments.** A brief review of some further developments and successes of Bohr's theory will illustrate the impact which it made on science.

(a) *Ellipticity of orbits.* It had indeed been most pleasing to find in Bohr's model an atom that had some similarity to the general structure of the solar system (even though the details and laws of motion are fundamentally different). While this analogy was

destined to disappear more and more in the subsequent developments, one of the first steps was the elaboration of Bohr's initial circular orbits to include other, elliptical orbits with large eccentricities. For example, the hydrogen atom in the excited state  $n = 2$  or  $3$  or  $4$  may have its electron in either circular or elliptic orbits, as shown in Fig. 25-11. The period of orbital revolution, and with it the angular momentum, are the same for all orbits drawn for the same quantum number  $n$ , because in each set half of the major axis of each of the ellipses has the same value as the radius of the corresponding circle. By Kepler's law, already assumed to apply to the motion of electrons in permitted orbits, the period of revolution  $T$  depends only on the semimajor axis  $R$  (see Section 9-4).

It would seem, therefore, that it should make no difference whether the electrons are assumed to have a circular or an elliptical orbit. However, in the latter case, they may come very close to the nucleus, and like a comet near the sun, move at the perihelion with great speed—so great that the mass there increases somewhat through the relativistic effect previously discussed. This produces a slight change of the orbit, a slow rotation or precession of the whole elliptical path once every 40,000 orbital revolutions, instead of a single fixed closed curve (Fig. 25-12), a precession that is also observed for comet paths. But there is also a small effect on the total energy in such an orbit, and if a transition of the electron occurs from one state to another, say  $n = 4$  to  $n = 3$ , the energy level differences are not quite the same for jumps to different ellipses.

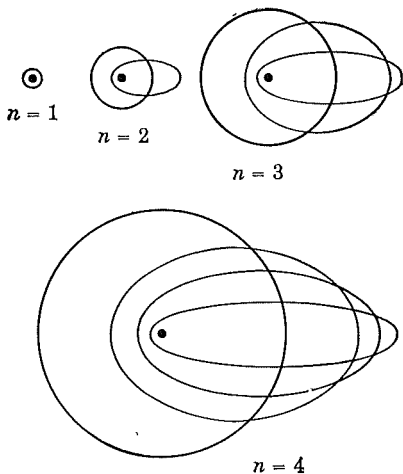


FIG. 25-11. Relative positions and dimensions of elliptical orbits corresponding to first four quantum numbers  $n$ . (A. Sommerfeld's modification for hydrogen.)

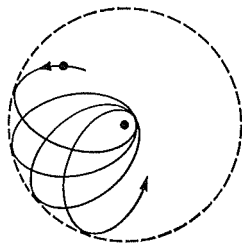


FIG. 25-12. Owing to the increase of mass near the nucleus, the electron's orbit is not a closed ellipse but a rosette figure.

But there is also a small effect on the total energy in such an orbit, and if a transition of the electron occurs from one state to another, say  $n = 4$  to  $n = 3$ , the energy level differences are not quite the same for jumps to different ellipses.

Consequently, most hydrogen lines should actually be made of several components. This "fine structure" had indeed been observed with very good spectroscopes since 1887, and it was understandably a great triumph both for the Bohr atom and for relativity theory that for the most part the calculated and observed fine structure of these spectral lines agreed.

(b) *Value of Rydberg's constant.* In an experimental field of research as precise as spectroscopy it was disconcerting to discover that the *empirical* values of Rydberg's constant  $R$  for hydrogen and the value *calculated* from Eq. (25-14), although very close, were not in precise agreement. Specifically, the observed wavelengths  $\lambda$  for hydrogen are all slightly higher than those given by Bohr's initial equation.

Part of the discrepancy was removed by Bohr in another widely acclaimed amendment to his original work. As you will recall, in our derivation of Eq. (25-14) there was the implicit assumption that the nucleus is fixed, that all orbital radii are to be measured to the nucleus, and that the latter has no KE. Actually however, the nucleus, not having an infinite mass compared with the electron, participates in the motion, just as the earth and moon rotate together about a common point about 1000 miles from the earth's center. This changes the calculated constant by a significant amount, bringing it into better agreement with the empirical value.

PROBLEM 25-17. (a) Following this qualitative line of reasoning, what should be the magnitude of the discrepancy for the spectrum of ionized helium? Is it larger or smaller than for hydrogen? (This prediction proved correct also.) (b) Attempt to describe how the participation of the nucleus in the motion will affect the quantitative correction of Eq. (25-14). (For exact solution, see texts such as H. Semat, *Atomic Physics* (1946), p. 191.)

PROBLEM 25-18. The wavelength of the line  $H_{\alpha}$  emitted by gas consisting of the heavy natural isotope of hydrogen,  ${}^2\text{H}_2$ , is about 1.8 Å shorter than for the lighter isotope. (This is, in fact, how the presence of this isotope was first confirmed in 1932.) Explain qualitatively the reason for this shift.

(c) *Line spectra of sodium.* It proved impossible to build up atoms with more than one orbital electron which would give a theoretical prediction of energy levels that fitted the observed facts of spectroscopy. In part, the difficulty was that a single electron moves in the constant field of the nucleus, so that the stability of orbits can be calculated with the aid of such simple classical formulas as Eq. (25-2), whereas in an atom with more than one electron each

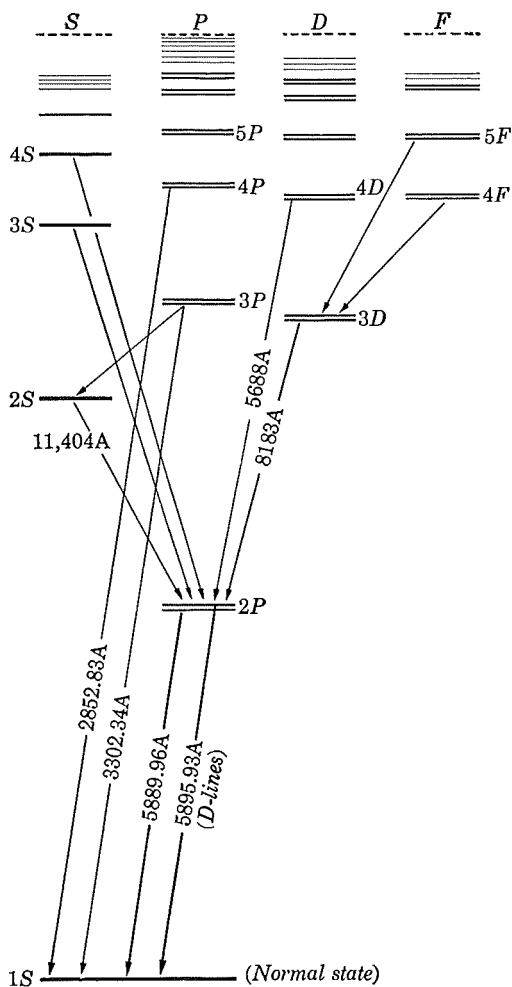


FIG. 25-13. Simplified energy level diagram for sodium. Only a few lines are indicated, and except for the D lines, only one member of each doublet.

finds itself in a rapidly fluctuating field owing to the motion of the others. Therefore the method was reversed, and the energy levels were derived from the known spectra, a procedure akin to the one we adopted initially in Section 25-2, although now much refined by the lessons learned from the hydrogen model.

Figure 25-13 gives the simplified energy level diagram for sodium which will serve as a typical example for other elements. It is to be imagined that there coexist in the atom several energy level scales, of which four (marked  $S$ ,  $P$ ,  $D$ ,  $F$ ) are shown here. Each observed photon from glowing sodium vapor can be explained as energy given up in transitions from a step in one ladder to one in a *neighboring* ladder. For example, the prominent yellow line in the visible spectrum (" $D$  line") corresponds to a transition from the level marked  $2P$  to  $1S$ , the latter being the normal state of the neutral, unexcited atom. The fact that this yellow line under higher resolution turns out to be a doublet of very close wavelengths is expressed on the diagram by making the level  $2P$  a double level; for similar reasons other levels also have such multiple structures.

One may visualize this energy level scheme roughly as follows. Each of the transitions refers to the activity of the *one* electron which is farthest away from the nucleus. The various levels within one vertical column differ from one another by the degree of excitement of the atom and the size of the outermost orbit, but the different columns represent varying degrees of eccentricity of the elliptic orbits in the excited state.

PROBLEM 25-19. An electron beam is allowed to pass into a tube containing a little sodium vapor (unexcited Na atoms). When these electrons are accelerated by 2.1 volts or more, but not before, the vapor emits yellow light of about  $5890 \times 10^{-8}$  cm wavelength. Explain the physical details of this phenomenon. What do you expect to happen to the light emission as the speed of the electron beam is gradually increased?

(d) *The shell structure of atoms.* As we have seen, Bohr visualized in 1913 that in the heavier elements the electrons formed a series of stable rings around the nucleus. This "flat" atom was first elaborated by the introduction of elliptic orbits. Figure 25-14 shows a set of representations which Bohr himself used in his lectures;\* the orbits are drawn approximately to scale. It is evident that the rare gases (He, Ne) have a close and tight structure, whereas

---

\* *Die Naturwissenschaften*, XI (1923), Table I.



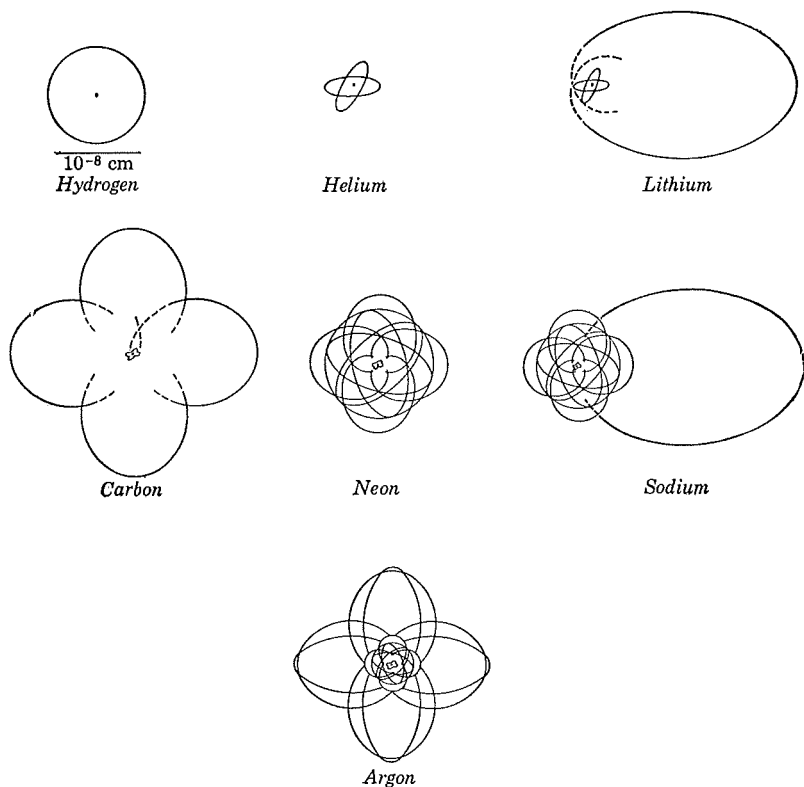


FIG. 25-14. Representations of the electronic orbits of some atoms, as prepared by Bohr for his own lectures.

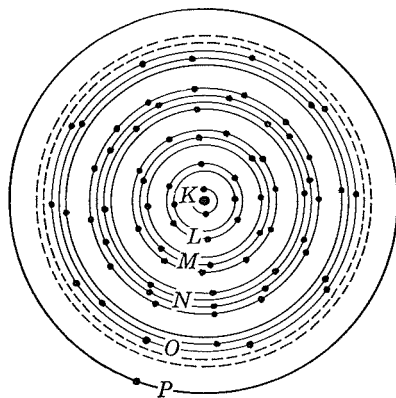


FIG. 25-15. Schematic representation of electrons in an atom of gold, showing the shells and subshells.

the alkali metals (Li, Na) have one easily detached electron in a far-flung orbit, corresponding to their valences of unity. Carbon has four loosely bound electrons, corresponding to its valence of 4.

In this pictorial scheme, intended mostly for symbolic purposes, the orbits were thought to trace out patterns in three dimensions. For example, the two orbits for He in Fig. 25-14 represent two circles at about  $60^\circ$  with respect to each other. Soon, however, these definite visualizable arrangements were replaced by a more abstract, highly useful concept of electron arrangements in atoms (G. N. Lewis, 1916). The neutral atom is thought to have its electrons distributed in shells around the nucleus. The innermost shell is called the *K* shell, and has room for a maximum of 2 electrons. Helium's two electrons "fill" or "close" the shell, so that lithium's third electron must start a new one, the *L* shell. The latter has room for 8 electrons, so that neon, with a total of 10 electrons, fills completely the *K* and *L* shells. The next element is sodium, whose 11th electron places itself in the 3rd or *M* shell; similarly for all other elements.

Altogether there are 7 main shells (*K,L,M,N,O,P,Q*), and each consists of as many parts as is indicated by its ordinal number on counting from the nucleus: *K*, the first shell, has no division; *L*, the second, consists of two subshells; *M*, the third, of three, and so forth. (See Fig. 25-15 for schematic diagram.) The first subshell in any shell can always hold up to 2 electrons, the 2nd up to 6, the third up to 10, the fourth up to 14, etc. And up to argon, the eighteenth element, the build-up of electrons proceeds quite simply (e.g., argon has two *K* electrons, eight *L*, then two in the 1st *M* subshell, six in the 2nd *M* subshell). But after that the electrons may be placed first into an outer shell before an inner one is filled.

Data for some of the elements are given in Table 25-1. We notice at once that some features correspond to Bohr's thought in 1913, e.g., the distribution of electrons in iron, cobalt, and nickel, where the number of electrons in the outermost shell stays constant, while the *M* shell gains one electron between steps. If the chemical behavior is governed, as here assumed, by the external electrons, we seem to be coming close to a physical explanation of chemical similarities.

This is also made evident by several other features of Table 25-1. The rare earths (lanthanum to lutecium), notoriously difficult to separate from one another in mixtures, are seen to be identical in the structure of the *four* outer subshells. The rare gases all have in

Table 25-1. Electron Configuration of the Elements

Shell		K	L	M	N	O	P	Q
At. No. (Z)	Element	(2)	(2) (6)	(2) (6) (10)	(2) (6) (10) (14)	(2) (6) (10) (14) (18)	(2) (6) (10) (14) (18) (22)	(2) (6) (10) (14) (18) (22) (26)
1	Hydrogen	1						
2	Helium	2						
3	Lithium	2	1					
6	Carbon	2	2					
8	Oxygen	2	2	4				
9	Fluorine	2	2	5				
10	Neon	2	2	6				
11	Sodium	2	2	6	1			
12	Magnesium	2	2	6	2			
17	Chlorine	2	2	6	2	5		
18	Argon	2	2	6	2	6		
19	Potassium	2	2	6	2	6	1	
20	Calcium	2	2	6	2	6	2	
21	Scandium	2	2	6	2	6	1	2
22	Titanium	2	2	6	2	6	2	2

Table 25-1 (Cont.)

Shell		K		L		M			N			O				P				Q				
Maximum number of electrons per subshell →		(2)	(2)	(2)	(6)	(2)	(6)	(10)	(2)	(6)	(10)	(14)	(2)	(6)	(10)	(14)	(18)	(2)	(6)	(10)	(14)	(18)	(22)	(2) (Etc.)
At. No. (Z)	Element																							
26	Iron	2	2	2	6	2	6	6	2															
27	Cobalt	2	2	2	6	2	6	7	2															
28	Nickel	2	2	2	6	2	6	8	2															
36	Krypton	2	2	2	6	2	6	10	2	6														
54	Xenon	2	2	2	6	2	6	10	2	6	10		2	6										
55	Cesium	2	2	2	6	2	6	10	2	6	10		2	6				1						
56	Barium	2	2	2	6	2	6	10	2	6	10		2	6				2						
57	Lanthanum	2	2	2	6	2	6	10	2	6	10		2	6	1			2						
58	Cerium	2	2	2	6	2	6	10	2	6	10	1	2	6	1			2						
59	Praseodymium	2	2	2	6	2	6	10	2	6	10	2	2	6	1			2						
71	Lutecium	2	2	2	6	2	6	10	2	6	10	14	2	6	1			2						
72	Hafnium	2	2	2	6	2	6	10	2	6	10	14	2	6	2			2						
86	Radon	2	2	2	6	2	6	10	2	6	10	14	2	6	10			2	6					
92	Uranium	2	2	2	6	2	6	10	2	6	10	14	2	6	10	3		2	6	1				2

common the idiosyncrasy that their electrons fill the first two subshells of their outermost shells (except, of course, for helium, which fills the *K* shell completely). Evidently there is a *particular stability* for this electron arrangement. Turning to the alkali metals (here we note lithium, sodium, and potassium), it is seen they each follow immediately after a rare gas and so have one lonely electron in the outermost shell, while the rest of the electrons are in a particularly stable configuration. On the other hand, the halogens, preceding a rare gas in each case, each lack one electron from filling the first two subshells.

(e) *The periodic table.* We have already begun to "explain" the periodic table in terms of electron configurations. The periodicity is governed by the recurrence of the rare gas arrangements. And while it is perfectly true that initially the details in Table 25-1 were so arranged that they would coincide with the known chemical periodicities, yet it was later possible, through the work of W. Pauli in 1925, to compile such a table with the aid of new fundamental physical principles rather than from a previous knowledge of detailed chemical behaviors.

A good deal of chemistry can be derived from such a table if one allows three postulates: that the chemical atom is indeed modeled on such a shell pattern, that the electron configurations are particularly stable in rare gas patterns (2, 2 + 8, 2 + 8 + 8, 2 + 8 + 18 + 8, etc.), and that atoms in molecular combinations are predisposed to redistribute electrons in patterns as near as possible to that of a stable rare gas. When, for example, hydrogen and lithium combine, one may imagine that the total of four electrons redistribute themselves around the nucleus as shown in Fig. 25-16, namely, into two helium-like patterns. This leaves the hydrogen atom negatively charged and the lithium atom positively charged, so that they will mutually attract each other by electric forces. In fact, these forces are so strong that such a

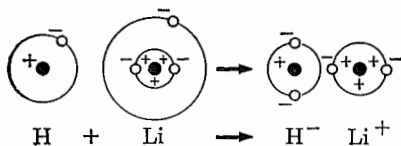


FIG. 25-16. Transfer of electron in the formation of electrovalent compound.

substance at ordinary temperatures forms a regular solid crystal lattice, alternately occupied by  $\text{H}^-$  and  $\text{Li}^+$  ions. Substances whose molecules are formed by the transfer of one or more electrons from one atom to another are called *ionic* or electrovalent substances.

PROBLEM 25-20. Many other substances are similarly explainable. Carry through such an argument, with diagrams, for NaCl and for NaF, which also form solids.

PROBLEM 25-21. On the basis of this conceptual scheme, what do we expect the volume per atom of a rare gas to be, compared with the next highest element in the periodic table? (Compare Fig. 19-7.) And what should we expect its relative ionization potential to be?

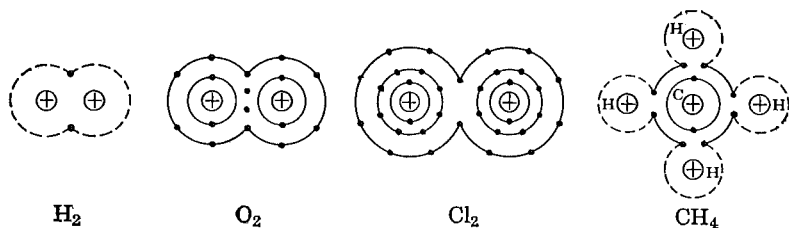


FIG. 25-17. Schematic representation of covalent compounds formed by electron sharing.

This scheme does not work for such molecules as  $\text{H}_2$ ,  $\text{O}_2$ ,  $\text{Cl}_2$ , where there can be no question about electron transfer because each partner in the diatomic molecule must show the same "affinity" for electrons as the other. We recall that this was precisely one of the most salient points against the early acceptance of Avogadro's molecular hypothesis. However, one may attempt a partial explanation by introducing the concept of "electron sharing." The two electron clouds join by sharing one or more electrons; for example, two of the total of 34 electrons of two Cl atoms are shared in the molecule  $\text{Cl}_2$ , four of the 16 in  $\text{O}_2$ , six of the 14 in  $\text{N}_2$ . In the organic compounds, the four outer electrons of carbon may be shared with hydrogen atoms, as in  $\text{CH}_4$  (Fig. 25-17). The tendency to produce rare-gas configurations is again affirmed.

All these cases of electron sharing are primarily examples of *covalent* substances, but there are a great many cases of chemical bonding between the extreme types of ionic and covalent substances. In the group called *polar* molecules, a subgroup of

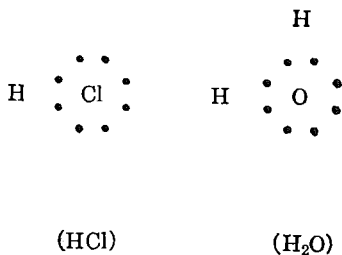


FIG. 25-18. Polar molecules. Only the outermost electrons are represented; note, however, that the shared electrons are asymmetrically placed.

covalent molecules, the shared electrons "stand" closer to one atom or the other, so that one part of the molecule is negative with respect to the other; this is illustrated in Fig. 25-18 for HCl and H<sub>2</sub>O, in the symbolism used in chemistry.

(f) *Valence*. The foregoing has hinted at the broad outlines of an electron theory of valence. The "hooks," bonds, or coupling mechanisms between atoms in compounds, to which reference was made in our previous discussion of valence (Section 19-8) begin to take on physical meaning. As we just saw, the simple cases of bonding are the transfer or the sharing of electrons to form around the nuclei such electron configurations as resemble most the nearest rare-gas structures. Thus magnesium and oxygen form MgO or, to use the structural formula, Mg=O, by the transfer of two of the 12 electrons from the magnesium atom to add to the oxygen atom's eight electrons, leaving each with its electrons in a pattern resembling neon atoms. The transfer of two electrons corresponds to the valence of 2 for each of these two elements. A glance at Table 25-1 explains why calcium and barium usually have a valence of 2, scandium 3, uranium 6.

Furthermore, it is evident why some elements, e.g., alkali metals, tend to form positive ions by losing electrons as in electrolysis or in forming electrovalent compounds; these are the so-called *electro-positive* elements, and their valence is conveniently identified as positive. On the other hand, nonmetals such as the halogens, oxygen, and sulphur, tend to form negative ions by gaining electrons to supplement their shells; they are designated as electronegative and their valence is identified as negative. The same terminology applies to radicals acting as a group, as for example to (OH) in NaOH. Hydrogen itself is unique; having only one electron to lose and needing only one to close the shell, it may show a positive or negative valence (of 1) depending on the role it plays in the compound. Carbon, nitrogen, and other elements, standing approximately equally far from rare-gas configurations, may also have positive or negative valences depending on whether they combine with metals or nonmetals.

PROBLEM 25-22. Decide on the electron configurations, valences, and, if possible, the type of bonding involved in each of these cases: MgS, CCl<sub>4</sub>, SO<sub>2</sub>, NH<sub>3</sub>, KCl, CaCl<sub>2</sub>, UF<sub>6</sub>. (Additional electron structures: S = 2; 2,6; 2,4. N = 2; 2,3.)

PROBLEM 25-23. If we had never seen this substance, why would we still expect the molecule CaO to exist?

PROBLEM 25-24. (a) What is the relationship between the structure of an element's outermost electron shell and the charge transferred per ion in electrolysis? (b) Why should metals be good conductors of electricity whereas nonmetals are not? (c) Why should there be any elements that have properties in common both with metals and with nonmetals? Where should one look for these elements in the periodic table? (d) How can we account for the fact that the valence of an element is different in different compounds?

PROBLEM 25-25. Draw diagrams representing the electron configurations for an atom of sodium, potassium, and cesium. What do you expect to be the order of chemical activity (violence of reaction with nonmetals like a halogen) in these three cases, and why?

When the periodic table is re-examined in the light of the theory of electron shells, one now finds a physical sanction for Mendeléeff's decision to arrange elements in order of increasing valence in each series. For example, in the second series of Table 19-3, the number of valence electrons (electrons beyond the shell corresponding to a rare-gas pattern) increases by one as we move from lithium at the left end of the row to carbon in the middle; thus the main valence increases in steps from  $+1$  to  $+4$ . At carbon one may say that the trend reverses; that is, now one approaches a full rare-gas pattern, and carbon may have the valence  $-4$ , nitrogen  $-3$ , oxygen  $-2$  and fluorine  $-1$ , each number corresponding to the "vacancies" in the pattern. A corresponding rise and fall in valence numbers exists in each period, as the elements first recede from and then approach a rare-gas configuration. Closer study of Table 25-1 reveals also a physical background for the structure of the periodic table of elements (Table 19-3): each period begins with an element whose outermost electron occupies an entirely new shell, and then gradually fills the vacancies in the subshells until at the end of each period a rare-gas configuration is achieved.

In summary, not only the main physical properties but also the chemical behavior of the elements seems to be governed in principle by the structure of the electron cloud around the nucleus. But the configuration itself seems determined largely by the atomic number, i.e., the *number* of electrons present, and this recognition, a triumph of old Pythagorean and neo-Platonic tendencies in modern science, has brought this remarkable reaction from Bohr himself:

"This interpretation of the atomic number may be said to signify an important step toward the resolution of a problem which for a long time has been one of the boldest dreams of nat-



ural science, namely, to build up an understanding of the regularities of nature upon the consideration of pure numbers."

(g) *Sharp lines vs. continuous spectra.* One of the initial puzzles concerning line spectra was the sharpness of the lines. Now we see clearly why an atom in a gas radiates at characteristic frequencies, uninfluenced by the fact that it may undergo many million collisions per second with other atoms. The average kinetic energy of thermal motion,  $\overline{KE}_{\text{trans}}$  per atom, is given by the kinetic theory as  $\overline{KE}_{\text{trans}} = \frac{3}{2} kT$ , where  $k$  = Boltzmann's constant. At room temperature, or about 300°K, this is about  $6 \times 10^{-14}$  erg. However, the energies needed to excite an atom in its normal state, the lowest ground state, are generally of the order of 10 to 100 times higher, as indicated by the fact that the lowest frequencies of line absorption are  $10^{14}$  to  $10^{15}$  sec<sup>-1</sup>. In short, the energy of thermal agitation is below the value of the least quantum which an atom in that state can accept, and therefore the interatomic collisions are perfectly elastic. Even when the temperature is high, the absorbed energy serves only to change the energy state of the atom in accordance with its particular energy level diagram.

PROBLEM 25-26. The lowest frequency, highest wavelength photon which a neutral lithium atom can absorb is  $\lambda = 6708$  Å. At what temperature will the  $\overline{KE}_{\text{trans}}$  of the colliding atoms correspond to this smallest quantum of energy accepted by an unexcited lithium atom?

However, there *are* effects of temperature on the emitted lines. For example, a radiating atom moving at high speeds toward or away from an observer, owing to a large random thermal motion at high temperatures, emits at a slightly larger or smaller frequency; this is due to the same Doppler effect that is so well known for sources of sounds (e.g., jet planes) that move with respect to the observer. For a large number of moving emitters, the net result is a *broadening* of the spectral lines, from which one may calculate the temperature of a distant source of the line spectrum.

The reason why glowing solids and liquids emit continuous spectra rather than line spectra is now clear also: the atoms are so close together that *no distinct energy level scheme can exist for the electron in the outermost orbit.* This condition applies also to a very hot gas, in which an electron is usually surrounded by several ionized atoms rather than being bound to one specific nuclear field.

(h) *The specific heat problem.*

We recall that the kinetic theory left us with an unresolved puzzle in the anomalous temperature dependence of the specific heat of molecular hydrogen gas (Section 20-5). The question was, why does the rotational degree of freedom begin to "freeze in" below about 70°K? Or to put it another way, why do thermal collisions of hydrogen gas molecules not begin to change their state of rotational motion until temperatures of above 40°K are reached?

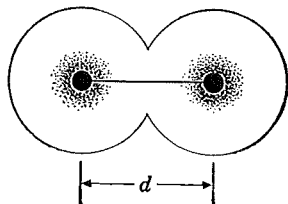


FIG. 25-19.

Consider a "dumbbell" molecule of  $H_2$ , and imagine it as in Fig. 25-19. The electron orbits are now of no interest, for all the mass is contained in the two nuclei, each having about  $1.6 \times 10^{-24}$  gm ( $m_H$ ). The distance  $d$  between them is about  $10^{-8}$  cm. Thus the rotational inertia  $I$  for rotation about the common center is given by  $2[m_H(d/2)^2]$  or  $8 \times 10^{-41}$  gm-cm<sup>2</sup>. The kinetic energy of rotation is  $\frac{1}{2} I \omega^2$ . But now we adopt the rule, transferred from Bohr's theory for electron orbits and found applicable throughout physics, that the angular momentum  $I\omega$  of a rotator is quantized according to  $I\omega = nh/2\pi$ , where  $n$  is an integer, namely, 1 for the ground state of the rotator, 2 for the first excited state, etc. Therefore we now must investigate whether the thermal collisions at low temperature have enough energy to put the molecule into the first state of rotation ( $n = 1$ ).

$$KE_{\text{rot}} = \frac{1}{2} I \omega^2 = \frac{1}{2} I \left( \frac{I\omega}{I} \right)^2 = \frac{n^2 h^2}{8\pi^2 I} = 7 \times 10^{-15} \text{ erg.} \quad (25-23)$$

Now, is this quantity of energy available at low temperature by a transfer of  $KE_{\text{trans}}$  into  $KE_{\text{rot}}$ ?

$$\overline{KE}_{\text{trans}} = \frac{3}{2} kT = 2.1 \times 10^{-16} T. \quad (25-24)$$

Not until  $T$  is greater than 33°K will the last expression give values equal to or higher than the  $7 \times 10^{-15}$  erg needed for the lowest state of rotation of the  $H_2$  molecule! We note that despite our rather crude approximations, the order of magnitude,  $> 33^\circ\text{K}$  from theory,  $> 40^\circ\text{K}$  from experiment, is in gratifying agreement. In summary: at low temperatures, the available average energy is not sufficient to provide the smallest acceptable quantum of energy for rotational motion.\*

\* The argument has been much simplified here.

**25-7 Epilogue.** The list of successes of Bohr's fundamental approach has by no means been exhausted in this recital. We might continue, as is done in the books recommended at the end of this chapter, to apply the theory and its subsequent modifications to solving such puzzles as the precise wavelengths of x-ray spectra (e.g., Moseley's unexplained  $K_\beta$  line), the difference in the appearance of line spectra under different experimental conditions, the relative intensities of some of the observed lines, and much besides.

But the fact is that Bohr's conceptual scheme, while today still indispensable and basic to an initial understanding of both physics and chemistry, has by now been overtaken by the relentless and swift progress of modern physical science. Several fundamental transformations became necessary to account for the large wealth of phenomena not included even in such a sweeping synthesis as Bohr's. In essence, these transformations centered on abandoning the last vestiges of visualizability of single events in the electron cloud; the rather comfortable and intuitively meaningful orbits, shells, and "jumps" of electrons had to be given up as essentially meaningless from an operational point of view, and, in fact, misleading for some purposes of prediction. The contemporary atom, which still retains the now familiar concepts of energy levels and quantization of angular momentum, quite intentionally no longer presents any simple physical picture to guide our imagination. It is a mathematical model, a set of postulates and equations—the ultimate triumph of ascetic thought. As Bohr himself said at the very beginning of this development (1925):

"To the physicists it will at first seem deplorable that in atomic problems we have apparently met with such a limitation of our usual means of visualization. This regret will, however, have to give way to thankfulness that mathematics in this field, too, presents us with the tools to prepare the way for further progress."

While the new results so far have been as gratifying as those of Bohr had been in their time, and do, in fact, account for all the successes of the older model in a larger context, they have not, however, solved every problem that Bohr's scheme left unanswered. Actually, the task was attacked from many sides; the prominent names include Heisenberg, Schrödinger, de Broglie, Born, Jordan and Dirac. We can hear the echo of every great innovator in this opening phrase from the publication by W. Heisenberg (1925) which initiated the theoretical revisions:

"In this paper I am going to attempt to find the foundation for a mechanics of quantum theory. This mechanics is based exclusively on relations between quantities which are observable in principle [i.e., frequencies and intensities of line spectra, and not electron orbits] . . ."

It is the old advice for a science in difficulties: Let us turn from our old preconceptions and start afresh from the phenomena. Just in this manner did Galileo say ". . . We, however, have decided to consider the phenomena of freely falling heavy bodies with an acceleration *such as actually occurs in nature* . . ."

The task is by no means finished. Will it ever be? Can it ever be? We who have followed the development of physical science from its infancy in ancient Greece to its flowering in our day, and who have seen that every Ptolemy is challenged by a Copernicus, that every Kepler and Galileo must be made meaningful by a Newton, that every Dalton is eventually transfigured by the work of a Mendeléeff or a Bohr, are fully prepared to accept the view of science as an ever-widening spiral, an endless quest for wider horizons. And knowing now in some measure both the main achievements of physical science and the task of the scientist as a person, we know also that as long as there is thought there can be no end to the adventures of the mind.

### Further References

M. Born, *The Restless Universe*. London: Blackie and Son (1935). Good reading, largely qualitative.

H. Buckley, *op. cit.* (Ch. 15). Chapters XIV and XV.

N. D. Cheronis *et al.*, *op. cit.* (Ch. 17). Chapters 28-33 review chemistry in terms of modern atomic structure.

J. C. Gregory, *op. cit.* (Ch. 18). Chapters XII through XV.

J. H. Hildebrand, *Principles of Chemistry*. New York: Macmillan (1947). Chapter XVII on the periodic system of the elements.

B. Jaffe, *op. cit.* (Ch. 19). On J. J. Thomson, Moseley.

L. B. Loeb and A. S. Adams, *op. cit.* (Ch. 20). Part VI, "The Electrical Structure of Matter, and the New Physics."

B. Russell, *The A B C of Atoms*. New York: E. P. Dutton (1923). A simplified account of atomic structure and quantum physics.

H. N. Russell *et al.*, *op. cit.* (Ch. 9), volume II. Applications of spectroscopy and atomic theory to astrophysics.

E. Zimmer, *The Revolution in Physics*. London: Faber and Faber (1936). Chapters IV to VIII. Good account of further developments.

## ANSWERS TO PROBLEMS

- 1-6. 7 mi/hr. 1-11. Almost 27 ft/sec. 1-13.  $3.6 \times 10^{-4}$  sec.
- 3-4. (a) 84 ft, 156 ft, -84 ft. (b) 68 ft/sec, 4 ft/sec, -124 ft/sec. (c) 9.4 sec. (d)  $-32$  ft/sec<sup>2</sup>. 3-6. 94 sec, 35,200 ft, 244,000 ft. 3-8. 3600 ft/sec. 3-10. 271 sec, 329 sec.
- 4-5. 26 gm, 0.26 cm/sec<sup>2</sup>. 4-6. 2.8 gm. 4-7.  $1.5 \times 10^7$  dynes. 4-9.  $9.809 \times 10^5$  dynes in Paris,  $9.801 \times 10^5$  dynes in Washington, 0.08%. 4-10. 0.5 dyne. 4-15. 24,700 dynes. 4-17. 294 cm/sec<sup>2</sup>, 12,700 dynes. 4-18. 1560 gm. 4-22.  $5.8 \times 10^6$  dynes.
- 5-1. 2450 radians or  $1.4 \times 10^5$  degrees of arc; 62 cm/sec and 244 cm/sec; 8.2 rad/sec. 5-3. 0.042 rev/hr, 0.289 mi/sec,  $7.3 \times 10^{-5}$  rad/sec. 5-4. 18.6 mi/sec. 5-8. (a) 12 mi/hr at 24° S of E. (b) 7.8 mi/hr at 79° S of E. 5-9.  $15\frac{1}{2}$  pdl at 11° toward  $F_2$  from  $F_1$ ;  $15\frac{1}{2}$  pdl in a direction exactly opposite to the resultant of  $F_1$  and  $F_2$ . 5-10. 2.9 cm/sec<sup>2</sup>, 0.56 cm/sec<sup>2</sup>. 5-14. 26 rad/sec. 5-16. 1.01 rad/sec<sup>2</sup>, 93 sec,  $2.5 \times 10^5$  degrees. 5-17. (a) 24.826 cm. (b) 2.17 cm. (c) Lose about 100 sec/day. 5-18. 977.5 cm/sec<sup>2</sup>. 5-20. (a)  $7.74 \times 10^7$  dynes. (b) 41.7 cm from the bridge.
- 6-2.  $10^2$  A.U. (order of magnitude);  $2.5 \times 10^5$  A.U.
- 11-5. 105 cm. 11-6. 1 dyne-cm<sup>2</sup>/tm<sup>2</sup>. 11-8.  $6 \times 10^{29}$  gm. 11-11. 5.5 gm/cm<sup>3</sup>, 3.7 gm/cm<sup>3</sup>, 1.7 gm/cm<sup>3</sup>, 172 cm/sec<sup>2</sup>, 31,600 cm/sec<sup>2</sup>. 11-12. 22,500 ft/sec; about 2 hr. 11-15. Slightly over 0.7 rev/hr.
- 16-1. (a) and (b), less than 2 sec. 16-2. 386 lb-ft/sec downward. 16-7. (a) 30 ft/sec. (b)  $2.3 \times 10^{-21}$  ft/sec. (c)  $5.62 \times 10^6$  pdl. (d)  $5.62 \times 10^6$  pdl. 16-11.  $v_{Kr}:v_{Ba} = 1.5:1$ . 16-16.  $2.25 \times 10^5$  gm-cm<sup>2</sup>,  $1.12 \times 10^5$  gm-cm<sup>2</sup>,  $0.90 \times 10^5$  gm-cm<sup>2</sup>. 16-18.  $-6.9 \times 10^{10}$  gm-cm<sup>2</sup>/sec,  $2.3 \times 10^7$  dynes. 16-19.  $1.7 \times 10^{35}$  lb-ft<sup>2</sup>/sec.
- 17-4. (a)  $5.28 \times 10^4$  ft-lbf,  $7.16 \times 10^{11}$  ergs,  $7.16 \times 10^4$  joules. (b)  $1.356 \times 10^7$  ergs. (c)  $-5.28 \times 10^4$  ft-lbf. (d) Zero. (e) Zero. 17-6.  $2.9 \times 10^8$  ergs. 17-7. 204 joules,  $\Delta KE = 13.1$  joules,  $\Delta PE = 147$  joules. 17-9. (a) 591 joules. (b) 291 joules. 17-12. (a) 99 cm/sec. (b)  $4.9 \times 10^5$  ergs. (c)  $PE = KE = 2.45 \times 10^6$  ergs. 17-29. 10 gm/day, 4400 kgm/yr. 17-34. 0.8°C. 17-36.  $7.07 \times 10^{40}$  gm-cm<sup>2</sup>/sec,  $0.05 \times 10^{27}$  gm,  $128 \times 10^{29}$  joules.
- 18-2. Volume = 0.239 liter.
- 19-6. 32.04, 46.07, 74.12, 324.41, 116.16, 258.10. 19-8. 25.8 tons. 19-9. 159.70; 84.03; 111.70; 132.03. 19-13.  $1.67 \times 10^{-24}$  gm;  $26.7 \times 10^{-24}$  gm;  $59.4 \times 10^{-24}$  gm.

20-1. (a)  $650^\circ\text{K}$ . 20-4.  $c_{\text{H}_2}:c_{\text{N}_2} = 3.75:1$ . 20-6. Slightly over  $6 \times 10^5$  cm/sec. 20-8. (a)  $5.9 \times 10^{10}$  ergs,  $0.623 \times 10^{10}$  ergs. (b)  $2.74 \times 10^{10}$  ergs. 20-9. (a)  $0.5 \times 10^{-48}$  gm-cm<sup>2</sup>. (b) Either  $0.5 \times 10^{-48}$  gm-cm<sup>2</sup> [Fig. 20-9(c)] or  $10^{-40}$  gm-cm<sup>2</sup> [Fig. 20-9(a) or (b)]. 20-10. At just over  $2000^\circ\text{C}$ ;  $\overline{\text{KE}}_{\text{trans}} = 1.1 \times 10^{-12}$  erg;  $C_v = 6.95$  (cal/mole)/deg.

21-10. (c)  $9 \times 10^{-3}$  dyne,  $5 \times 10^{-42}$  dyne. (e)  $2.1 \times 10^{10}$  electrons. 21-14. 12 dynes,  $1.33 \times 10^{-5}$  coul. 21-22. 0.096 erg/statcoul, 0.072 erg/statcoul. 21-25.  $3.4 \times 10^{-4}$  coul. 21-26.  $3.2 \times 10^{-6}$  erg,  $2.9 \times 10^{10}$  cm/sec (after making relativistic correction). 21-27. About  $1.6 \times 10^{-12}$  erg or  $1.6 \times 10^{-10}$  joule.

22-1.  $2.07 \times 10^{-5}$  cm,  $1.16 \times 10^{-4}$  cm,  $10^{-3}$  cm. 22-6. (b) 26 lines. 22-8. 972 Å, 913 Å.

23-2. (a)  $\text{PE}_{\text{max}} = 800$  ergs,  $\text{PE}_{\text{equil}} = \text{zero}$ ,  $\text{KE}_{\text{max}} = \text{zero}$ ,  $\text{KE}_{\text{equil}} = 800$  ergs, total energy = 800 ergs,  $f =$  slightly below 1 vib/sec. (b) 1000 ergs, 600 ergs, 4.5 cm, 3.5 cm. 23-4.  $0.43 \times 10^{15}$  sec<sup>-1</sup>,  $2.84 \times 10^{-12}$  erg,  $4.96 \times 10^{-12}$  erg. 23-6.  $1.5 \times 10^{-29}$  cm. 23-7.  $1.1 \times 10^8$  cm/sec. 23-12. (a)  $1.2 \times 10^{10}$  cm/sec. (b)  $8 \times 10^{-8}$  erg/photon. 23-14. (b)  $10.2 \times 10^8$  cm/sec, at  $44^\circ$  with the direction of the incident photon.

24-6. (a)  $2.66 \times 10^{21}$ . (b) Slightly over  $8.3 \times 10^{19}$ . 24-15.  $2.7 \times 10^{-12}$  cm. 24-18. (a)  $2.48 \times 10^4$  volts. 24-22. Binding energy of  ${}^9_4\text{Be}$ ,  $9.3 \times 10^{-5}$  erg;  ${}^{20}_{10}\text{Ne}$ ,  $25.7 \times 10^{-5}$  erg.

25-11.  $6.563 \times 10^{-5}$  cm and  $1.21 \times 10^{-5}$  cm (also original radiation at  $1.026 \times 10^{-5}$  cm). 25-14.  $4 \times 10^4$  °K. 25-15.  $16 \times 10^{-8}$  cm. 25-26.  $1.43 \times 10^4$  °K.

## APPENDIX I

### FUNDAMENTAL CONSTANTS\*

NAME OF QUANTITY	SYMBOL	VALUE
Velocity of light	$c$	$2.9979 \times 10^{10}$ cm/sec
Charge on electron	$e$	(-) $4.802 \times 10^{-10}$ statcoul
Rest mass of electron	$m_e$	$9.107 \times 10^{-28}$ gm
Planck's constant	$h$	$6.624 \times 10^{-27}$ erg-sec
Boltzmann constant	$k$	$1.380 \times 10^{-16}$ erg/degree K
Avogadro's number	$N$	$6.023 \times 10^{23}$ molecules/mole
Universal gas constant	$R$	$8.314 \times 10^7$ erg/mole-degree K
Mass of neutral hydrogen ( $H^1$ )		1.008128 amu (phys. scale)**
Mass of proton ( $p$ )		1.00758 amu
Mass of neutron ( $n$ )		1.00896 amu
Ratio of proton to electron mass		1836.14
1 amu		$1.660 \times 10^{-24}$ gm
Rydberg constant for infinitely heavy nucleus	$R_\infty$	$109,737$ cm <sup>-1</sup>
Rydberg constant for hydrogen	$R_H$	$109,678$ cm <sup>-1</sup>
Gravitational acceleration (standard value)	$g$	$980.665$ cm/sec <sup>2</sup>
Universal gravitational con- stant	$G$	$6.67 \times 10^{-8}$ dyne-cm <sup>2</sup> /gm <sup>2</sup>
Mass of earth		$5.98 \times 10^{27}$ gm
Mean radius of earth		$6.37 \times 10^8$ cm
Mean distance from earth to sun		$9.29 \times 10^7$ miles = $1.49 \times 10^{13}$ cm
Eccentricity of earth's orbit		0.017
Mean distance from earth to moon		$3.84 \times 10^{10}$ cm
Sun's diameter		$1.39 \times 10^{11}$ cm

\* The atomic constants are based largely on the work of J. W. M. DuMond and E. R. Cohen, *Phys. Rev.* **82**, 555 (1951). In each case, the figures are known to higher accuracy than implied here, but are given in the form most useful for the purposes of this book. For further data, consult a collection such as the *Handbook of Chemistry and Physics* (Chemical Rubber Publishing Co., Cleveland; issued yearly).

\*\* More recently it has been suggested that the correct values are: for  $H^1$ , 1.008142; for  $p$ , 1.007593; for  $n$ , 1.008982.

Sun's mass		$333,000 \times$ mass of earth
Mechanical equivalent of heat	$J$	4.185 joule/cal
Standard atmospheric pressure		$1.013 \times 10^6$ dynes/cm <sup>2</sup>
Volume of perfect gas at 1 atm, 0°C		22.415 liter/mole
Faraday constant (1 faraday)	$F$	96,500 coul/mole
Numerical constants: $\pi = 3.142$ , $e = 2.718$ , $\sqrt{2} = 1.414$ , $\sqrt{3} = 1.732$ .		



## APPENDIX II

### TABLE OF CONVERSION FACTORS

#### LENGTH:

- 1 m = 100 cm = 1000 mm
- 1 km = 1000 m = 0.6214 mile
- 1 m = 39.37 in.; 1 cm = 0.3937 in.
- 1 ft = 30.48 cm; 1 in. = 2.540 cm
- 1 mi = 5,280 ft = 1.609 km
- 1 Å =  $10^{-8}$  cm;  $1\mu$  (micron) =  $10^{-4}$  cm
- 1 A.U. =  $9.287 \times 10^7$  miles =  $14.95 \times 10^7$  km

#### AREA:

- 1 cm<sup>2</sup> = 0.155 in.<sup>2</sup>; 1 m<sup>2</sup> = 10<sup>4</sup> cm<sup>2</sup> = 10.76 ft<sup>2</sup>
- 1 in.<sup>2</sup> = 6.452 cm<sup>2</sup>; 1 ft<sup>2</sup> = 144 in.<sup>2</sup> = 0.0929 m<sup>2</sup>

#### VOLUME:

- 1 liter = 1000 cm<sup>3</sup> = 10<sup>-3</sup> m<sup>3</sup> = 0.0351 ft<sup>3</sup> = 61 in.<sup>3</sup>
- 1 ft<sup>3</sup> = 0.0283 m<sup>3</sup> = 28.32 liters; 1 in.<sup>3</sup> = 16.39 cm<sup>3</sup>

#### VELOCITY:

- 1 cm/sec = 0.03281 ft/sec; 1 ft/sec = 30.48 cm/sec
- 1 mile/min = 60 mi/hr = 88 ft/sec

#### ACCELERATION:

- 1 cm/sec<sup>2</sup> = 0.03281 ft/sec<sup>2</sup> = 0.01 m/sec<sup>2</sup>
- 30.48 cm/sec<sup>2</sup> = 1 ft/sec<sup>2</sup> = 0.3048 m/sec<sup>2</sup>
- 100 cm/sec<sup>2</sup> = 3.281 ft/sec<sup>2</sup> = 1 m/sec<sup>2</sup>

#### FORCE:

- 1 dyne =  $7.23 \times 10^{-5}$  pdl =  $2.247 \times 10^{-6}$  lbf =  $10^{-5}$  newton
- $1.383 \times 10^4$  dynes = 1 pdl = 0.0311 lbf = 0.1383 newton
- $4.45 \times 10^5$  dynes = 32.2 pdl = 1 lbf = 4.45 newtons
- 10<sup>5</sup> dynes = 7.23 pdl = 0.2247 lbf = 1 newton

#### MASS:

- 1 gm =  $2.205 \times 10^{-3}$  lb =  $6.85 \times 10^{-5}$  slug =  $10^{-3}$  kgm
- 453.6 gm = 1 lb = 0.0311 slug = 0.4536 kgm
- $1.459 \times 10^4$  gm = 32.2 lb = 1 slug = 14.59 kgm
- 10<sup>3</sup> gm = 2.205 lb = 0.0685 slug = 1 kgm

#### PRESSURE:

- 1 atm = 14.7 lb/in.<sup>2</sup> =  $1.013 \times 10^6$  dynes/cm<sup>2</sup>

## ENERGY:

1 joule =  $10^7$  ergs = 0.239 cal; 1 cal = 4.18 joule

1 ev =  $10^{-6}$  Mev =  $1.60 \times 10^{-12}$  erg =  $1.07 \times 10^{-9}$  amu

1 amu =  $1.66 \times 10^{-24}$  gm =  $1.49 \times 10^{-3}$  erg = 931 Mev

## ELECTRIC UNITS:

1 statcoul =  $3.33 \times 10^{-10}$  coul; 1 coul =  $3 \times 10^9$  statcoul

1 statvolt = 300 volts; 1 volt =  $3.33 \times 10^{-3}$  statvolt

## Greek Letter Symbols

A	$\alpha$	Alpha	N	$\nu$	Nu
B	$\beta$	Beta	$\Xi$	$\xi$	Xi
$\Gamma$	$\gamma$	Gamma	O	$o$	Omicron
$\Delta$	$\delta$	Delta	$\Pi$	$\pi$	Pi
E	$\epsilon$	Epsilon	P	$\rho$	Rho
Z	$\zeta$	Zeta	$\Sigma$	$\sigma$	Sigma
H	$\eta$	Eta	T	$\tau$	Tau
$\Theta$	$\theta$	Theta	Y	$\upsilon$	Upsilon
I	$\iota$	Iota	$\Phi$	$\phi$	Phi
K	$\kappa$	Kappa	X	$\chi$	Chi
$\Lambda$	$\lambda$	Lambda	$\Psi$	$\psi$	Psi
M	$\mu$	Mu	$\Omega$	$\omega$	Omega

## APPENDIX III

## ALPHABETIC LIST OF THE ELEMENTS

ELEMENT	SYMBOL	ATOMIC	ELEMENT	SYMBOL	ATOMIC
		NUMBER			NUMBER
		Z			Z
Actinium	Ac	89	Holmium	Ho	67
Aluminum	Al	13	Hydrogen	H	1
Americium	Am	95	Indium	In	49
Antimony	Sb	51	Iodine	I	53
Argon	A	18	Iridium	Ir	77
Arsenic	As	33	Iron	Fe	26
Astatine	At	85	Krypton	Kr	36
Barium	Ba	56	Lanthanum	La	57
Berkelium	Bk	97	Lead	Pb	82
Beryllium	Be	4	Lithium	Li	3
Bismuth	Bi	83	Lutetium	Lu	71
Boron	B	5	Magnesium	Mg	12
Bromine	Br	35	Manganese	Mn	25
Cadmium	Cd	48	Mercury	Hg	80
Calcium	Ca	20	Molybdenum	Mo	42
Californium	Cf	98	Neodymium	Nd	60
Carbon	C	6	Neon	Ne	10
Cerium	Ce	58	Neptunium	Np	93
Cesium	Cs	55	Nickel	Ni	28
Chlorine	Cl	17	Niobium	Nb	41
Chromium	Cr	24	Nitrogen	N	7
Cobalt	Co	27	Osmium	Os	76
Copper	Cu	29	Oxygen	O	8
Curium	Cm	96	Palladium	Pd	46
Dysprosium	Dy	66	Phosphorus	P	15
Erbium	Er	68	Platinum	Pt	78
Europium	Eu	63	Plutonium	Pu	94
Fluorine	F	9	Polonium	Po	84
Francium	Fr	87	Potassium	K	19
Gadolinium	Gd	64	Praseodymium	Pr	59
Gallium	Ga	31	Promethium	Pm	61
Germanium	Ge	32	Protactinium	Pa	91
Gold	Au	79	Radium	Ra	88
Hafnium	Hf	72	Radon	Rn	86
Helium	He	2	Rhenium	Re	75

ALPHABETIC LIST OF THE ELEMENTS—*Continued*

ELEMENT	SYMBOL	ATOMIC	ELEMENT	SYMBOL	ATOMIC
		NUMBER			NUMBER
		Z			Z
Rhodium	Rh	45	Thallium	Tl	81
Rubidium	Rb	37	Thorium	Th	90
Ruthenium	Ru	44	Thulium	Tm	69
Samarium	Sm	62	Tin	Sn	50
Scandium	Sc	21	Titanium	Ti	22
Selenium	Se	34	Tungsten		
Silicon	Si	14	(Wolfram)	W	74
Silver	Ag	47	Uranium	U	92
Sodium	Na	11	Vanadium	V	23
Strontium	Sr	38	Xenon	Xe	54
Sulphur	S	16	Ytterbium	Yb	70
Tantalum	Ta	73	Yttrium	Y	39
Technetium	Tc	43	Zinc	Zn	30
Tellurium	Te	52	Zirconium	Zr	40
Terbium	Tb	65			

APPENDIX IV

PERIODIC TABLE OF THE ELEMENTS

Atomic weights are the most recent adopted by the International Union of Chemistry; none are given for artificially produced elements. The full names of the elements are given in Appendix III.

GROUP →	I	II	III	IV	V	VI	VII	VIII	O	
Period	Series									
1	1 H 1.0080								2 He 4.003	
2	3 Li 6.940	4 Be 9.013	5 B 10.82	6 C 12.010	7 N 14.008	8 O 16.000	9 F 19.00		10 Ne 20.183	
3	11 Na 22.997	12 Mg 24.32	13 Al 26.97	14 Si 28.06	15 P 30.98	16 S 32.066	17 Cl 35.457		18 Ar 39.944	
4	19 K 39.096	20 Ca 40.08	21 Sc 45.10	22 Ti 47.90	23 V 50.95	24 Cr 52.01	25 Mn 54.93	26 Fe 55.85	27 Co 58.94	28 Ni 58.69
5	29 Cu 63.54	30 Zn 65.38	31 Ga 69.72	32 Ge 72.00	33 As 74.91	34 Se 78.96	35 Br 79.916		36 Kr 83.7	
6	37 Rb 85.48	38 Sr 87.63	39 Y 88.92	40 Zr 91.22	41 Nb 92.91	42 Mo 95.95	43 Tc	44 Ru 101.7	45 Rh 102.91	46 Pd 106.7
7	47 Ag 107.880	48 Cd 112.41	49 In 114.76	50 Sn 118.70	51 Sb 121.76	52 Te 127.61	53 I 126.92		54 Xe 131.3	
8	55 Cs 132.91	56 Ba 137.36	57-71 Rare earths*	72 Hf 178.6	73 Ta 180.88	74 W 183.92	75 Re 186.31	76 Os 190.2	77 Ir 193.1	78 Pt 195.23
9	79 Au 197.2	80 Hg 200.61	81 Tl 204.39	82 Pb 207.21	83 Bi 209.00	84 Po 210	85 At		86 Rn 222	
10	87 Fr	88 Ra 226.05	89 Actinide series**							

\* Rare earths: 57 La | 58 Ce | 59 Pr | 60 Nd | 61 Pm | 62 Sm | 63 Eu | 64 Gd | 65 Tb | 66 Dy | 67 Ho | 68 Er | 69 Tm | 70 Yb | 71 Lu | 138.92 | 140.13 | 140.92 | 144.27 | 144.27 | 150.43 | 152.0 | 156.9 | 159.2 | 162.46 | 164.94 | 167.2 | 169.4 | 173.04 | 174.99

\*\* Actinide series: 89 Ac | 90 Th | 91 Pa | 92 U | 93 Np | 94 Pu | 95 Am | 96 Cm | 97 Bk | 98 Cf | 227 | 232.12 | 231 | 238.07 | 238.07



## APPENDIX V

### SUMMARY OF SOME TRIGONOMETRIC RELATIONS

- (1) **Right triangles:** (a) The sum of the angles in any plane triangle being  $180^\circ$ , and angle  $\gamma$  in Fig. A-1 being  $90^\circ$ , it follows that  $\alpha + \beta = 90^\circ$ .  
 (b) By the Pythagorean theorem,  $c^2 = a^2 + b^2$ .
- (2) **Definitions:** We can *define* six functions of any one angle in a right triangle, namely sine (*sin*), cosine (*cos*), tangent (*tan*), cotangent (*cot*), secant (*sec*), and cosecant (*cosec*). Specifically,

$$\sin \alpha = \frac{\text{opposite side}}{\text{hypotenuse}} = \frac{a}{c}, \quad \cos \alpha = \frac{\text{adjacent side}}{\text{hypotenuse}} = \frac{b}{c},$$

$$\tan \alpha = \frac{a}{b}, \quad \cot \alpha = \frac{b}{a}, \quad \sec \alpha = \frac{c}{b}, \quad \text{cosec } \alpha = \frac{c}{a}.$$

(For the first three functions, these ratios are given in Appendix VI.) It follows directly from these definitions:

$$(a) \quad \tan \alpha = \frac{\sin \alpha}{\cos \alpha}, \quad \cot \alpha = \frac{\cos \alpha}{\sin \alpha} = \frac{1}{\tan \alpha}, \quad \sec \alpha = \frac{1}{\cos \alpha},$$

$$\text{cosec } \alpha = \frac{1}{\sin \alpha}.$$

Therefore only the sine and cosine functions of an angle need be discussed in detail.

$$(b) \quad \sin^2 \alpha = (\sin \alpha)^2 = \frac{a^2}{c^2},$$

$$\cos^2 \alpha = (\cos \alpha)^2 = \frac{b^2}{c^2}.$$

Thus,  $\sin^2 \alpha + \cos^2 \alpha = \frac{a^2 + b^2}{c^2} = 1$  (by the Pythagorean theorem).

- (c) Because  $\sin \beta = \frac{b}{c}$ ,  $\cos \beta = \frac{a}{c}$  and so on, therefore

$$\sin \alpha = \cos \beta, \quad \cos \alpha = \sin \beta,$$

$$\sin \alpha = \cos (90 - \alpha), \quad \cos \alpha = \sin (90 - \alpha).$$

**EXAMPLE 1.** In a right triangle ( $\gamma = 90^\circ$ ),  $\alpha = 35^\circ$ , length  $b = 4$  cm; what is length  $a$ ?

$$\tan \alpha = \frac{a}{b}, \quad \text{therefore } a = b \tan \alpha = 4 \times 0.700 = 2.8 \text{ cm.}$$

**EXAMPLE 2.** In a right triangle,  $a = 8.3$  cm,  $c = 21$  cm; what are  $\alpha$  and  $\beta$ ?

$$\sin \alpha = \frac{a}{c} = \frac{8.3}{21} = 0.395.$$

As seen in the table, this lies between  $\sin 23^\circ = 0.391$  and  $\sin 24^\circ = 0.407$ . But

$$\frac{0.395 - 0.391}{0.407 - 0.391} = \frac{0.004}{0.016} = \frac{1}{4}.$$

Thus angle  $\alpha$  is (roughly)  $\frac{1}{4}$  of  $1^\circ$  higher than  $23^\circ$ , or  $23^\circ 15'$ .  
And  $\beta = 90^\circ - \alpha = 66^\circ 45'$ .

- (3) One can define the same trigonometric functions for any angle, even for an angle not in a triangle but enclosed between an inclined line and the horizontal (Fig. A-2). If  $r$  is the length of the line and  $(x, y)$  the  $x$  and  $y$  coordinates of its end point  $P$ , the other end being at the origin of the coordinate system, then

$$\sin \theta = \frac{y}{r}, \quad \cos \theta = \frac{x}{r}, \quad \tan \theta = \frac{y}{x}.$$

By convention,  $\theta$  is measured counterclockwise from the right (positive) section of the abscissa.

- (4) If  $\theta$  is larger than  $90^\circ$ , its trigonometric functions are defined as before, e.g.,  $\sin \theta = \frac{y}{r}$ .

- (a)  $\theta = 90^\circ$  to  $180^\circ$  (Fig. A-3). Note that  $x$  has a negative value, being on the left side of the  $x$ - $y$  coordinate system, or in the second quadrant. Thus

$$\cos \theta = \frac{x}{r} = -\cos \varphi = -\cos (180^\circ - \theta),$$

$$\sin \theta = \frac{y}{r} = \sin \varphi = \sin (180^\circ - \theta).$$

- (b)  $\theta = 180^\circ$  to  $270^\circ$  (Fig. A-4);  $\varphi = \theta - 180^\circ$ . Both  $x$  and  $y$  have negative values

$$\sin \theta = -\sin \varphi = -\sin (\theta - 180^\circ),$$

$$\cos \theta = -\cos \varphi = -\cos (\theta - 180^\circ).$$

- (c)  $\theta = 270^\circ$  to  $360^\circ$  (Fig. A-5);  $\varphi = 360^\circ - \theta$ . Now  $y$  has a negative value,  $x$  a positive one:

$$\sin \theta = -\sin \varphi = -\sin (360^\circ - \theta),$$

$$\cos \theta = \cos \varphi = \cos (360^\circ - \theta).$$

These relations show that trigonometric tables computed for the range  $0^\circ$  to  $90^\circ$  will serve for all angles.

EXAMPLE 3. To find  $\sin \theta$  and  $\tan \theta$  if  $\theta = 295^\circ$ . Note that  $\theta$  is between  $270^\circ$  and  $360^\circ$ ;  $\sin \theta = -\sin (360^\circ - \theta) = -\sin 65^\circ = -0.906$ . And  $\tan \theta = \frac{\sin \theta}{\cos \theta} = \frac{-0.906}{0.423} = -2.15$ .



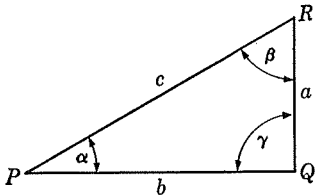


Fig. A-1

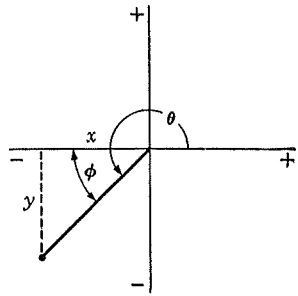


Fig. A-4

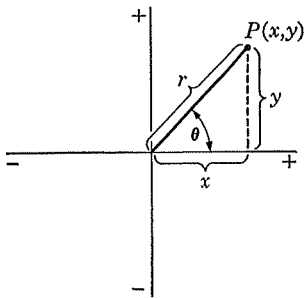


Fig. A-2

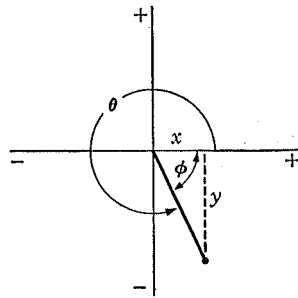


Fig. A-5

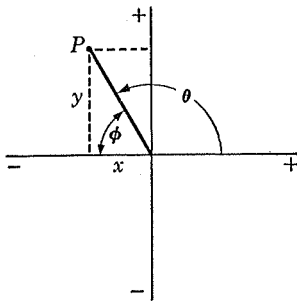


Fig. A-3

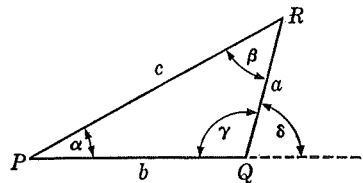


Fig. A-6

(5) **General triangles:** Even when a triangle does not have a right angle, as in Fig. A-6, there still exist several simple relations between the sides and the angles.

(a) **Law of Cosines:** The square of any one side is equal to the sum of the squares of the other two sides minus twice the product of those two into the cosine of their included angle. For example,

$$c^2 = a^2 + b^2 - 2 ab \cos \gamma.$$

Thus one may find one side in terms of two others and one angle, or one angle in terms of three sides. Because the complementary angle  $\delta = 180 - \gamma$ , and  $\cos \delta = \cos (180 - \gamma) = -\cos \gamma$ , we may write

$$c^2 = a^2 + b^2 + 2 ab \cos \delta.$$

(b) **Law of Sines:** In any triangle, the ratio of any side and the sine of the opposite angle is constant, or

$$\frac{a}{\sin \alpha} = \frac{b}{\sin \beta} = \frac{c}{\sin \gamma}.$$

Thus, two sides and one angle determine the remaining angles and side, and so forth.

(6) **Other useful trigonometric relations:**

$$\sin (\alpha + \beta) = \sin \alpha \cos \beta + \cos \alpha \sin \beta,$$

$$\cos (\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta,$$

$$2 \sin \alpha \cos \alpha = \sin 2\alpha.$$

$$\cos 2\alpha = \cos^2 \alpha - \sin^2 \alpha.$$

# APPENDIX VI

## NATURAL TRIGONOMETRIC FUNCTIONS

Angle		Sine	Cosine	Tan- gent	Angle		Sine	Cosine	Tan- gent
Degree	Radian				Degree	Radian			
0°	.000	0.000	1.000	0.000	46°	0.803	0.719	0.695	1.036
1°	.018	.018	1.000	.018	47°	.820	.731	.682	1.072
2°	.035	.035	0.999	.035	48°	.838	.743	.669	1.111
3°	.052	.052	.999	.052	49°	.855	.755	.656	1.150
4°	.070	.070	.998	.070	50°	.873	.766	.643	1.192
5°	.087	.087	.996	.088	51°	.890	.777	.629	1.235
6°	.105	.105	.995	.105	52°	.908	.788	.616	1.280
7°	.122	.122	.993	.123	53°	.925	.799	.602	1.327
8°	.140	.139	.990	.141	54°	.942	.809	.588	1.376
9°	.157	.156	.988	.158	55°	.960	.819	.574	1.428
10°	.175	.174	.985	.176	56°	.977	.829	.559	1.483
11°	.192	.191	.982	.194	57°	.995	.839	.545	1.540
12°	.209	.208	.978	.213	58°	1.012	.848	.530	1.600
13°	.227	.225	.974	.231	59°	1.030	.857	.515	1.664
14°	.244	.242	.970	.249	60°	1.047	.866	.500	1.732
15°	.262	.259	.966	.268	61°	1.065	.875	.485	1.804
16°	.279	.276	.961	.287	62°	1.082	.883	.470	1.881
17°	.297	.292	.956	.306	63°	1.100	.891	.454	1.963
18°	.314	.309	.951	.325	64°	1.117	.899	.438	2.050
19°	.332	.326	.946	.344	65°	1.134	.906	.423	2.145
20°	.349	.342	.940	.364	66°	1.152	.914	.407	2.246
21°	.367	.358	.934	.384	67°	1.169	.921	.391	2.356
22°	.384	.375	.927	.404	68°	1.187	.927	.375	2.475
23°	.401	.391	.921	.425	69°	1.204	.934	.358	2.605
24°	.419	.407	.914	.445	70°	1.222	.940	.342	2.747
25°	.436	.423	.906	.466	71°	1.239	.946	.326	2.904
26°	.454	.438	.899	.488	72°	1.257	.951	.309	3.078
27°	.471	.454	.891	.510	73°	1.274	.956	.292	3.271
28°	.489	.470	.883	.532	74°	1.292	.961	.276	3.487
29°	.506	.485	.875	.554	75°	1.309	.966	.259	3.732
30°	.524	.500	.866	.577	76°	1.326	.970	.242	4.011
31°	.541	.515	.857	.601	77°	1.344	.974	.225	4.331
32°	.559	.530	.848	.625	78°	1.361	.978	.208	4.705
33°	.576	.545	.839	.649	79°	1.379	.982	.191	5.145
34°	.593	.559	.829	.675	80°	1.396	.985	.174	5.671
35°	.611	.574	.819	.700	81°	1.414	.988	.156	6.314
36°	.628	.588	.809	.727	82°	1.431	.990	.139	7.115
37°	.646	.602	.799	.754	83°	1.449	.993	.122	8.144
38°	.663	.616	.788	.781	84°	1.466	.995	.105	9.514
39°	.681	.629	.777	.810	85°	1.484	.996	.087	11.43
40°	.698	.643	.766	.839	86°	1.501	.998	.070	14.30
41°	.716	.656	.755	.869	87°	1.518	.999	.052	19.08
42°	.733	.669	.743	.900	88°	1.536	.999	.035	28.64
43°	.751	.682	.731	.933	89°	1.553	1.000	.018	57.29
44°	.768	.695	.719	.966	90°	1.571	1.000	.000	∞
45°	.785	.707	.707	1.000					

N	0	1	2	3	4	5	6	7	8	9
0	0000	0000	3010	4771	6021	6990	7782	8451	9031	9542
1	0000	0414	0792	1139	1461	1761	2041	2304	2553	2788
2	3010	3222	3424	3617	3802	3979	4150	4314	4472	4624
3	4771	4914	5051	5185	5315	5441	5563	5682	5798	5911
4	6021	6128	6232	6335	6435	6532	6628	6721	6812	6902
5	6990	7076	7160	7243	7324	7404	7482	7559	7634	7709
6	7782	7853	7924	7993	8062	8129	8195	8261	8325	8388
7	8451	8513	8573	8633	8692	8751	8808	8865	8921	8976
8	9031	9085	9138	9191	9243	9294	9345	9395	9445	9494
9	9542	9590	9638	9685	9731	9777	9823	9868	9912	9956
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
N	0	1	2	3	4	5	6	7	8	9

N	0	1	2	3	4	5	6	7	8	9
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996
100	0000	0004	0009	0013	0017	0022	0026	0030	0035	0039
N	0	1	2	3	4	5	6	7	8	9



## INDEX

- Abscissa, 1  
 Acceleration, angular, 98  
     centripetal, 93  
     definition of, 10  
     in free fall, 28 ff.  
     gravitational, 66, 185, 192-193  
     uniform, 26  
 Actinide series, 427  
 Adams, *John C.*, 195, 199  
 Adiabatic process, 459  
 Almagest, 123  
 Ampere, 498  
 Ångström, *A. J.*, 511, 523  
 Angstrom, definition of, 511  
 Angular momentum, definition of, 307  
     quantization of, 591, 599, 605, 621, 622  
 Aphelion, 154  
 Aquinas, *St. Thomas*, 127, 134, 175  
 Archimedes, 335  
 Aristarchus of Samos, 118-121, 129-130, 132  
 Aristotle, 18 ff., 368, 374  
     on astronomy, 116-118, 134-135  
     elements, 376  
 Asteroids, 198  
 Astronomy, in ancient Greece, 113-126  
 Atmosphere, escape of gases from, 461-462  
     mixture of gases in, 461  
 Atomic mass unit, definition of, 297  
 Atomic number, 427, 555, 568, 573-575  
 Atomic weight, 393-397, 408, 416-430, 499, 575  
 Atomism, ancient view, 367, 374, 376  
 Atwood's Machine, 74, 75, 280  
 Avogadro, *Amedeo*, 383, 400-406  
 Avogadro's hypothesis, 269, 411, 417, 434, 437, 438  
 Avogadro's number, 411, 437, 499  
  
*Bacon, Francis*, 136, 141, 148, 219, 227, 230, 245, 259, 368  
 Balance, equal arm, 67  
     spring, 59, 60  
 Ballistic pendulum, 329  
*Balmer, Johann Jakob*, 523-526, 528, 582, 588  
 Balmer series, 588, 589, 594, 596, 598, 601  
 Barometer, 370  
*Becquerel, Henri*, 255, 552  
*Benedetti*, 18, 263  
*Bernoulli, D.*, 375-376, 448  
*Berzelius, J. J.*, 392, 416  
*Bessel, F. W.*, 121  
*Beveridge, W. I. B.*, 257  
 Beta rays, origin of, 606, 607  
*Bethe, H. A.*, 343  
 Binding energy of nucleus, 340, 579  
*Black, Joseph*, 377-378, 380  
 Block-and-tackle, 335  
 Bode's Law, 196-202, 242, 362, 427  
*Bohr, Niels*, 146, 216, 265, 471, 570, 582-614, 619  
*Boltzmann, L.*, 434  
 Boltzmann's constant, definition of, 449  
*Born, Max*, 547, 622  
*Boyle, Robert*, 169, 370, 371  
     on chemistry, 383, 385  
     law of gases, 228, 370, 372, 373, 375, 434, 447, 460  
 Brackett series, 525, 594-595  
*Bridgman, P. W.*, 146, 225-226, 258, 263  
*Broek, A. van den*, 568  
*Brown, Robert*, 435  
 Brownian movement, 435, 436, 441, 453  
*Bruno, Giordano*, 148, 166, 202  
*Bunsen, R. W.*, 517, 522  
*Buridan, Jean*, 287  
*Butterfield, H.*, 134

- Calendar, 141  
 Caloric theory, 140, 284, 376-384,  
 388-390, 507  
 Calorie, definition of, 346  
 Calorimetry, 377-378  
*Cannizzaro, Stanislao*, 383, 406, 417  
*Cardinal Barberini*, 165  
*Carnap, Rudolph*, 230, 263  
 Cathode rays, 500, 552, 572, 574  
*Cavendish, Henry*, 184  
 Celestial sphere, 113, 161-162  
 Centripetal, acceleration, 93, 192  
 force, 95-97, 175-181, 203, 314  
*Chadwick, James*, 577  
*Charles, Jacques*, 371, 372  
 Chemical compounds, Dalton's the-  
 ory of, 390-398  
 Chemical elements, families of, 416-  
 418, 420, 425, 429, 616  
 Chemical equation, balancing of, 409  
 Circular motion, equation of, 98  
 uniform, 86 ff.  
 uniformly accelerated, 97  
*Claudius Ptolemy of Alexandria*,  
 planetary system of, 121-125, 129,  
 132, 135, 139-140, 161-164  
*Clausius, R.*, 434, 464  
 Cloud chamber, 296, 570, 575  
*Cohen, Morris*, 263  
 Coefficient of friction, 73  
 Cohesion, 440  
 Colors, wavelengths of, 512  
 Combustion, 282  
 Comets, 191, 504, 608  
 Component, of motion of projectile,  
 37  
*Compton, A. H.*, 303  
 Compton effect, 303, 550  
*Conant, James B.*, 167, 230, 248  
 Conductor, 477  
 Conservation of angular momentum,  
 law of, 308-312, 361  
 Conservation of charge, law of, 474  
 Conservation of energy, law of, 325-  
 364, 550  
 Conservation of heat, law of, 377,  
 381  
 Conservation of matter, law of, 279-  
 284, 341, 378, 391  
 Conservation of momentum, law of,  
 286-312, 314, 345, 549-550  
 Conservation of vis viva, law of, 315-  
 318, 329  
 Copernican system, 161-167  
*Copernicus, Nicolaus*, 124, 139, 148,  
 264  
 heliocentric theory, 127-137, 149  
 Corona of sun, 598  
 Correspondence principle, 590-593,  
 596  
 Cosmic rays, 299, 481  
 Coulomb, definition of, 484  
 Coulomb's law, 484-486, 567, 569,  
 585  
 Covalent compounds, 617  
*Curie, Irène*, 577  
*Curie, Pierre and Marie*, 255  
 Current, electric, 477  
 Cyclotron, 322  
  
*Dalton, John*, 140, 354, 365, 383, 387-  
 401, 418, 432, 448, 460, 461, 497,  
 550  
*Dante*, 125  
*Darwin, Charles*, 139, 220  
*Davy, Humphry*, 382, 383, 392, 432,  
 497  
*De Broglie, L. V.*, 622  
 Deferent, 121-122, 130  
 Definite proportions, law of, 393  
 Degree of freedom of motion, 464-  
 468, 522, 621  
*Democritus*, 279, 367  
*Descartes, René*, 33, 178, 219, 368, 377  
 on conservation of momentum, 291  
 vortex theory, 207-208, 309, 373-  
 374  
 Deuteron, 322  
 Deuterium, 556  
*Dewey, John*, 241, 263, 264  
 Diffusion, 441, 450  
*Dirac, P. A. M.*, 622  
 Discovery, process of, 220, 258  
*Döbereiner, Johann Wolfgang*, 417  
 Doppler effect, 620  
 Duality, wave-photon, 545-547  
*Dulong, P. L.*, 349



- Dumas, J. B. A.*, 416  
 Dyne, 61
- Eccentricity, definition of, 153  
 Eccentric motion, 121-122, 130, 152  
 Eclipse, 141  
*Eddington, Arthur*, 244, 583  
 Effusion, 450  
*Einstein, Albert*, 69, 216, 239, 264, 336, 340  
   photoelectric effect, 536, 540-546  
   photon theory, 583, 584  
   relativity theory, 142, 231, 506  
 Electric charge, definition of, 473-474  
   quantity of, 482-484  
 Electric current, 496-497  
   field strength, 487  
   intensity, 487  
   potential, 490-496  
 Electrolysis, 497-498  
 Electrolyte, 479  
 Electron, 475-476  
   charge of, 499-500  
   orbits in hydrogen atom, 583-599  
   sharing, 617  
   shell structure, 611-616  
 Electron-positron "annihilation," 340  
 Electron volt, 496, 558  
 Electroscope, 480  
 Electrostatic field, 486-490  
 Electrostatics, 473-496  
 Electrovalent compounds, 616  
 Elements, Aristotelian, 19  
   chemical, definition of, 385-386  
 Elevator problem, 75  
*Eliot, T. S.*, 216  
 Ellipse, 152-154, 176-177  
 Energy, definition of, 323  
   levels, 531-536, 583, 584-603, 611, 620  
   quantization of, 531-536  
 Energy-mass equivalence, 338-343  
*Epicurus*, 279, 367  
 Epicycle, 121-122, 124, 130, 152  
 Equant, 123, 129-130  
 Equations of motion, accelerated circular, 98-99  
   for constant acceleration, 11, 12, 13  
   projectiles, 47  
   Equilibrant, 93  
   Equinoxes, precession of, 202  
   Equipartition of energy, principle of, 466-467, 522, 534, 535  
   Equivalence of heat and work, 348-349  
   Erg, definition of, 320  
   Ether, 206-207, 377, 501-507  
   *Euclid*, 175  
   *Eudoxus*, 116  
   Expansion, thermal, 371-372  
   Extrapolation, 2
- Faraday, Michael*, 265, 349, 482, 497, 498, 500  
 Faraday (quantity of electric charge), 498  
*Fermi, Enrico*, 344  
 Field, electric, 486-490  
   electromagnetic, 501-511  
 Fluorescence, 547-548  
 Force, Aristotelian view, 58  
   centripetal, 95-97  
   definition of, 57  
   restoring, 103  
*r'oucault, J. B. L.*, 504  
 Frame of reference, for Newton's laws, 81  
*Franklin, Benjamin*, 473  
*Frank, P.*, 263  
*Fraunhofer, Joseph*, 517-521  
 Free fall, 16 ff.  
   law of, 228, 239  
 Frequency, rotation, definition of, 85  
*Fresnel, Augustine*, 503
- Galaxy, 202-204, 252  
*Galileo Galilei*, 17 ff., 220, 280, 368  
   on astronomy, 160-167  
   biographical sketch, 23-25, 165-166  
   free fall, 242, 265, 268  
   and Inquisition, 25  
   mathematical law, 227  
   influence on Newton, 189  
   pile driver problem, 318  
   primary and secondary quantities, 226  
   on projectile motion, 50ff., 138, 150, 262, 263, 270

- Galileo's *Two New Sciences*, 25 ff.
- Gas, equation of state, 438  
 general law of, 370-373  
 models of, 373-376, 387-390, 436-443  
 universal constant, 438
- Gassendi, P.*, 368
- Geiger, H.*, 564, 565, 567, 571
- Geiger counter, 546, 559
- Geocentric system, 161
- Gibbs, J. W.*, 434
- Gilbert, William*, 148
- Graaff, Van de*, 496
- Graham, Thomas*, 451
- Gram, 70
- Gram-atomic weight, definition of, 410
- Gram-molecular weight, definition of, 410
- Gravitation, constant of, 182-185  
 law of, 175-210
- Growth of science, 248-250
- Guericke, Otto von*, 370
- Gay-Lussac, J. L.*, 350, 352, 371, 440  
 law of combining volumes, 398-404, 434
- Hahn, Otto*, 144
- Halley, Edmond*, 170, 172, 178, 190, 191, 202
- Harmonic oscillator, 102
- Heat energy, 346  
 mechanical equivalent of, 349
- Heisenberg, W.*, 622
- Helium, line spectrum of, 599-600
- Helmholtz, Hermann von*, 356, 357, 434
- Heracles of Pontus*, 118
- Herapath, W.*, 432
- Herschel, William*, 195, 196, 198, 202, 504
- Hertz, Heinrich*, 508-511, 530, 537, 583
- Hipparchus of Rhodes*, 121
- Hooke, Robert*, 172, 189
- Horsepower, definition of, 348
- Human body, energy balance, 359
- Huygens, Christian*, 169, 189, 313, 314, 329
- on centripetal acceleration, 93, 172  
 theory of light, 503-504
- Idealism, 237
- Impulse, definition of, 307
- Inclined plane, 31-32
- Induction, electric, 478-481
- Inertia, 59  
 law of, 33
- Infeld, Leopold*, 69
- Inquisition, 25, 148, 165-166
- Insulators, 477
- Interpolation, 2
- Ion, 498, 553  
 definition of, 475
- Ionic compounds, 616
- Ionization potential, 597-598, 603
- Isomer, 413
- Isotopes, 552-562, 579, 607  
 separation of, 451-453, 553-554
- Joliot, Frédéric*, 577
- Jordan, P.*, 622
- Joule, James Prescott*, 277, 382  
 kinetic theory of heat, 432-435, 448, 463  
 on mechanical equivalent of heat, 353-356
- Joule, definition of, 321
- Jupiter, equatorial bulge, 191  
 moons of, 161-162, 164
- Kekulé, F. A.*, 220
- Kelvin, Lord*, see William Thompson
- Kepler, Johannes*, 148-161, 178, 186, 189, 191, 197, 207, 217, 301, 303, 515  
 first law, 150-153, 177, 183  
 second law, 154-156, 175-176, 183, 311, 312  
 third law, 156-157, 171, 177-178, 183, 187, 205, 310, 586, 608
- Kilogram, 70
- Kinematics, 18
- Kinetic energy, definition of, 322  
 of rotation, 331
- Kirchhoff, G. R.*, 517-520, 522
- Koyré, Alexandre*, 21
- Krönig, A.*, 434

- Lagrange, J. C.*, 325, 349  
 Latent heat of melting, 346  
*Lavoisier, Antoine Laurent*, 19, 281-284, 349, 378, 386-387, 393  
 LCAMm, see Conservation of angular momentum, law of  
 LCE, see Conservation of energy, law of  
 LCM, see Conservation of matter, law of  
 LCMm, see Conservation of momentum, law of  
*Leibniz, Gottfried Wilhelm*, 315, 368  
*Lenard, P.*, 552, 562  
 Length, definition of, 223  
     standard of, 62  
     units of, 2  
*Leonardo da Vinci*, 286  
*Leucippus*, 279, 367  
 Lever, 333-335  
*Leverrier, J. J.*, 196, 199  
*Lewis, G. N.*, 613  
 Light, continuous emission spectrum, 511-515, 528  
     infrared, 511-512  
     line absorption spectrum, 517, 601-602  
     line emission spectrum, 515-519, 575  
     momentum associated with, 302-303  
     speed of, 253  
     theory of, 501-511  
     ultraviolet, 511-512  
 Lines of force, 489-490  
*Lodge, Oliver*, 165, 194  
*Lowell, Percival*, 196  
*Lucretius*, 279, 367-368, 373, 401  
*Luther, Martin*, 134  
*Lyman, Theodore*, 594  
 Lyman series, 525-526, 594-596, 601  
  
*Mach, Ernst*, 83, 263, 267  
 Machines, 333-336, 358  
*Marconi, G.*, 510  
*Marcus Marci*, 291  
*Margenau, H.*, 263  
*Marsden, E.*, 564, 565, 567, 571  
 Mass, gravitational, 68, 69  
     inertial, 68, 69  
     relativistic increase of, 336-339  
     standard of, 60, 62, 82  
     units of, 2  
 Mass defect, 579  
 Mass-energy relation, 338-343, 549  
 Mass number, 555  
     definition of, 342  
 Mass spectrograph, 553, 554, 569  
 Mathematics, role in science, 227..  
     233, 268  
*Maxwell, James Clerk*, 434, 442, 457, 501-511, 540, 590  
*Mayer, Julius Robert*, 350-357, 382, 432  
 Mean free path, 439, 460  
*Mendel, Gregor*, 551  
*Mendeléeff, Dmitri Ivanovitch*, 418-430, 619  
 Mercury, precession of, 209  
 Mesons, 299  
 Metaphysics, 114-115, 244-245  
*Meyer, Julius Lothar*, 418, 429  
*Michelson, A. A.*, 506  
 Milky Way, 113, 161, 202-204  
*Millikan, Robert A.*, 499, 544  
*Milton, John*, 135, 236  
 Mole, definition of, 410  
 Molecular weights, table of, 407  
 Molecule, definition of, 401  
     dimensions of, 460  
     size of, 439  
     speed of, 453-459  
 Momentum, definition of, 287-288  
 Moon, Newton's work on its motion, 172-173, 179-181, 190  
*Morley, E. W.*, 506  
*Moseley, H. G. J.*, 571, 605, 606, 622  
 Multiple proportion, law of, 395  
  
*Nagaoka*, 562, 570  
*Nagel, E.*, 263  
 "Natural motion," 19  
 "Natural place," 19  
 Neptune, discovery of, 196, 199, 302  
 Neutrino, 344-345, 577  
 Neutron, 576  
     discovery of, 577

- Newlands, J. A. R.*, 417, 424  
*Newton, Isaac*, 33, 55, 264  
 on atoms, 368-369  
 biographical note, 171-173, 189  
 centripetal force, 95-96  
 definition of mass, 280  
 definition of momentum, 288  
 first law, 57, 280, 287  
 law of gases, 374, 387, 391  
 law of inertia, 57, 280, 287  
 laws of motion, 57 ff., 174  
 nonoperational concepts, 224  
 "Rules of Reasoning in Philosophy," 174, 181, 185  
 second law, 59-62, 72-78, 288  
 theistic assumption, 241-242  
 theory of light, 503-505, 540  
 theory of universal gravitation, 137, 175-210, 486  
 third law, 78-83  
 on tides, 193-194  
*Northrop, F. S. C.*, 263  
*Nostradamus*, 134  
 Nuclear energy, 255  
 Nuclear fission, 144, 296, 343  
 Nuclear fusion, 343, 360  
 Nucleus, discovery of, 564  
 size of, 569-570  
 structure of, 576-580  
 transmutation of, 342, 390
- Ockham, William*, 286-287  
 Operational definition, 9, 222, 235  
 Ordinate, 2  
 Organic compounds, structural formulas, 415  
 Oscillator, 102, 501, 508-511, 520-522, 530-536, 547, 583  
 Oxidation, 282
- Packing fraction, 580  
 Pair-production, 339  
 Parabola, 39, 48, 49, 231  
*Paracelsus*, 134, 385  
 Parallax, 119, 121, 125  
 Paschen series, 525, 589, 594-595  
*Pasteur, Louis*, 245  
*Pauli, Wolfgang*, 243, 616  
*Pearson, K.*, 263
- Pendulum, 102-104, 185, 257, 314, 327-329  
 Galilean, 328  
 inelastic collision, 293  
 isochronism of, 104  
 period of, 104  
 Perihelion, 154  
 Period of rotation, 100  
 definition of, 86  
 Periodic system of elements, 418-430, 604, 616  
 Perpetuum mobile, 358, 360  
 Perturbations, 186, 190, 201  
 Pfund series, 525, 594-595  
 Phlogiston theory, 140, 282, 283, 393  
 Photoelectric effect, 537-549, 601, 602  
 threshold frequency, 539, 543  
 work function, 543, 545  
 Photon, 542-551, 574, 583, 584, 587  
 Photosynthesis, 359-360  
*Pickering, E. C.*, 526  
*Pickering, W. H.*, 196  
 Pisa, leaning tower of, 18, 23-24, 26  
*Planck, Max*, 145, 259, 528-537, 583, 584, 590, 596  
 Planck's constant, 534-535, 544  
 Planetary system, geocentric, 116-118, 121-125  
 heliocentric, 118-121  
*Plato*, 114-115, 139, 261, 368  
 Pluto, discovery of, 196, 199  
*Poincaré, Henri*, 223, 243, 259, 263  
 Polar molecules, 617  
 Positivism, 237  
 Postulates of impotency, 358, 390, 506  
 Potential, electric, 490-496  
 Potential energy, definition of, 323  
 Pound, avoirdupois, 70  
 force, 71  
 Poundal, 70  
 Precession, of electron orbit, 608  
 of equinoxes, 202  
 of mercury, 209  
 Pressure, atmospheric, 369  
 gauge, 373  
*Priestley, Joseph*, 219, 257, 258  
 Probability, 458-459

- Probable error, 1, 223  
 Projectile motion, 26, 36 ff.  
 Proton, 322, 576  
*Prout, J. L.*, 393  
*Prout, William*, 415-416, 578  
 Pulley, 335  
*Pythagoras*, 113, 619
- Quantum number, 592  
 Quantum theory of light, 531-551
- Radian, 86  
 Radioactive series, 559-562  
 Radioactivity, discovery of, 255, 552  
   displacement law of, 561  
   half life, 556-558  
   law of decay, 232-233, 557  
 Range, 38  
 Rare earths, 427, 604, 613  
 Reaction-car experiment, 82, 104, 294  
 Realism, 237  
 Recoil of gun, 294-295  
 Reduction, chemical, 282  
 Reference circle, 100  
*Reichenbach, H.*, 263  
 Relativity, Galilean, 51-52  
   theory of, 506-507  
 Resonance, 520-522, 548  
 Rest energy, 341  
 Rest mass, 546, 549, 579  
 Resultant, 90, 91  
 Retrograde motion, 122  
*Richer, Jean*, 108  
*Robins, Benjamin*, 329  
*Röntgen, Wilhelm*, 265, 548  
 Rotational inertia, definition of, 305  
   of disk, 306  
   of hoop, 305  
   of particle, 305  
   of sphere, 306  
 Royal Society, 169, 540  
*Rumford, Count*, see Benjamin Thompson  
*Russell, Bertrand*, 172, 244, 263, 267, 476  
*Rutherford, Ernest*, 171, 342, 556, 563-570, 582, 583, 607
- Rydberg, J. R.*, 243, 525, 526, 596  
 Rydberg's constant, 609
- Sarton, George*, 219, 249, 367  
 Scattering of alpha-particles, 563-570  
*Schleiden, M. J.*, 551  
*Schrödinger, E.*, 622  
*Schwann, T.*, 551  
 Science, growth of, 248-250  
 Scientific journals, 251  
 Scientific method, 218-221, 257-273  
 Scintillation, 568, 575  
*Séguin, Marc*, 352, 356  
 Sidereal period, 141, 156  
 Simple harmonic motion, 99-106  
*Singer, Charles*, 281  
*Sizzi, Francesco*, 164  
 Slug, 71  
*Smyth, H. D.*, 236, 451  
*Soddy, Frederick*, 556, 559  
 Sodium, line spectrum of, 609-611  
 Solar system, evolution of, 360-363  
*Sommerfeld, A.*, 243  
 Sound waves, 109  
 Space, absolute, 173, 174, 224  
 Specific heat, 462-468, 621  
   at constant pressure, 347  
   at constant volume, 347  
   definition of, 346  
 Spectroscopy, 517-519, 522, 524, 609  
 Spectrum of electromagnetic radiation, table of, 548  
 Speed, average, 6  
   definition of, 3  
   instantaneous, 9  
*Spinoza, B.*, 175  
 Spring, vibration of, 109  
*Stevinus*, 18  
 Statcoulomb, definition of, 484  
 Standard of length, 253  
 Statics, principles of, 334  
*Stern, Otto*, 454  
*Stoney, G. Johnstone*, 500  
*Strassman, F.*, 144  
 Super-nova, 361  
 Superposition, law of, 44  
   principle of, 41  
 System, conservative, 325  
   dissipative, 325

- System of units, British Absolute, 70  
 British Engineering, 70  
 cgs absolute, 69  
 mks, 70  
 table, 71
- Tartaglia*, 18, 263  
*Taylor, F. S.*, 20  
 Temperature, absolute zero of, 372  
 Tension, 75  
 Theory, purpose of, 138-141  
 Thermionic tube, 538  
 Thermodynamics, laws of, 358  
*Thompson, Benjamin* (Count Rumford), 348, 352-354, 379-384, 387, 432  
*Thomson, William* (Lord Kelvin), 140, 234, 355, 372  
*Thomson, J. J.*, 537, 538, 553, 562, 564, 582, 602-603, 605  
 Thought experiments, 30, 60, 262  
 Time, absolute, 224  
 definition of, 252  
 units of, 2  
 Torque, definition of, 304  
*Torricelli, Evangelista*, 369-370, 373  
 Trajectory, 37, 48  
 Transmutation of nucleus, 342, 390  
*Tycho Brahe*, 148-149, 156, 189  
*Tyndall, John*, 353
- Ultracentrifuge, 95  
 Universal gravitation, constant of, 499  
 law of, 530  
 Uranus, discovery of, 195, 198
- Vacuum pump, 370  
 Valence, 412-418, 430, 575, 604, 613, 618  
*Van de Graaff, R. J.*, 496  
 Vectors, 42  
 addition of, 89, 90, 91  
 algebra of, 88-93  
 composition of, 42, 109, 174  
 graphic method of addition of, 44  
 resolution of, 42, 109  
 subtraction of, 89, 92  
*Velikovskiy, Immanuel*, 363-364  
 Velocity, angular, 87  
 instantaneous, 7, 8, 9  
 uniform, 26  
 Venus, phases of, 141, 161, 163  
 Vibration, amplitude of, 100, 106  
 frequency of, 107  
 period of, 102-106  
 string, 104-106  
*Vis viva*, definition of, 315  
*Volta, Alessandro*, 496
- Waterston, J. J.*, 434  
*Watt, James*, 348  
 Wave, electromagnetic, 501-511  
 propagation of, 106  
 sound, 109  
 Wavelength, definition of, 107, 507  
 Weight, 63 ff.  
*Weyl, H.*, 263  
*Whitehead, Alfred North*, 115, 143, 216, 240, 263, 280  
*Whittaker, Edmund*, 287  
 Wien's displacement law, 513-514, 533, 534  
*Wilson, C. T. R.*, 296  
*Winckler, C.*, 427  
 Work, definition of, 320  
 Work function, 549
- X-ray, 548-549, 605  
 discovery of, 265  
 spectra of, 571-575
- Young, Thomas*, 503, 545
- Zeeman, P.*, 510