C. O'D. ISELIN

A Survey Report On

HUMAN FACTORS

IN

UNDERSEA WARFARE

Prepared by the

PANEL ON PSYCHOLOGY AND
PHYSIOLOGY

COMMITTEE ON UNDERSEA WARFARE
NATIONAL RESEARCH COUNCIL
WASHINGTON, D. C. 1949

This volume is one of a series of survey reports prepared for the Office of Naval Research by technical advisory panels of The National Research Council Committee on Undersea Warfare.

Publication has been authorized under contract N7 orr-291, Task III between the Office of Naval Research and The National Academy of Sciences. All inquiries concerning this report should be addressed to the Committee on Undersea Warfare, National Research Council, Washington, D. C. or to the Chief of Naval Research, Navy Department, Washington, D. C.

PREFACE

The Committee on Undersea Warfare of the National Research Council was established in the fall of 1946 at the request of the Office of Naval Research. One of the principal functions of the committee was to advise the Navy upon matters pertaining to scientific problems and research related to submarine warfare. In considering these problems, the committee recognized that many of them are inextricably associated with the human factor, that is, with the adaptability and efficiency of the human individuals who will man the submarine and operate its technical equipment.

A Panel on Psychology and Physiology was appointed by the Committee and was authorized to draw up an outline for a survey of basic and applied research on problems related to the human factor in undersea warfare. On 8 July, 1947 this panel met at The Johns Hopkins University and outlined plans for the Survey. The principal topics and sub-topics to be covered in the survey report were decided upon and the authors were selected.

Despite the fact that many of the authors were busy in post-war academic and research activities, and some were still in the process of transition from wartime research to peacetime pursuits, there was almost one hundred percent acceptance of the Committee's invitation to prepare chapters for this volume. All of the authors are experienced investigators who have conducted research upon some phases of the field in which they were asked to write. Although a number of these men had served in research capacities in military or civilian agencies during the war, only a few had been directly connected with the submarine problem. This was not believed to be a serious handicap however, since it was felt that a fresh and somewhat detached point of view might prove valuable in pointing up new problems.

The purposes of this volume as outlined to each author were as follows: first, to provide a concise, but comprehensive, account of the present status of basic and applied research related to the efficiency of the human individual who will use the technical equipment and who will live under conditions such as are encountered in a submarine; and secondly, to formulate recommendations for future research, based on the usual scientific objective of increasing our knowledge of relevant phenomena.

Each author was informed that he need not consider in detail the problems associated with the specific items of equipment and the particular operations entailed. That is, he need not concern himself alone with a survey of research on each of the operations as they actually exist. Rather, he should emphasize the basic knowledge which may be expected to result from scientific research on the topics which encompass the fields of activity in which the submariner engages and the general conditions under which he works.

For example, there are numerous activities which involve the use of vision in particular ways and under special conditions and requirements. There are problems of instrument legibility which include the placement, type, and function of dials and indicators so that the maximum information may be obtained in a minimum of time and with the greatest accuracy. Other types of visual functions include the use of optical scopes, radar and sonar presentations, maps and charts, and lookout activities. In each case there is the problem of whether the task imposed by the equipment and the amount of information to be obtained from it is within the range of the capacities of the human individuals using it, whether they can translate the information, interpret it if necessary, and pass it on to control centers where it can be used to best advantage. It is obvious, therefore, that the design and arrangement of equipment and of the panel layouts must be adapted to the ease, comfort, and general efficiency of the operator. Similar problems occur in the auditory field, and both visual and sound communication add further problems of this nature.

Basic to the performance of any task are the physiological and psychological conditions under which a person must work. Consequently, habitability, or the conditions of temperature, humidity, pressure, diet, sleep, rest, and the effect of noise, color, lighting, and motion all enter into the general problem of over-all efficiency. The numerous factors which enter into physiological stress are of course extremely important. If one is fatigued, distracted, emotionally disturbed, anxious and tense, and generally over-wrought, efficiency is likely to suffer. Likewise, the effect of one's basic motivations, the morale and leadership problems, and general social relationships on board ship may seriously hamper performance of individuals and the whole crew.

Thus, it may be seen that in addition to selection and training problems, there are many other general considerations involving the physiological and psychological state of the individual which must be taken into consideration in seeking the optimal efficiency of the submarine crew. It is necessary to know what the environment on board a submarine is like, what the activities are, and how the submariner lives and works, but it should be strongly emphasized that the survey is not concerned with the specific details of these tasks and conditions. Rather, it is concerned with basic knowledge derived from research in these areas which may help solve the problems and which may point the way to problems on which research should be undertaken. This research may not be based on submarine problems *per se* but upon particular variables isolated for study in the laboratory. In still other instances the research may have been done in connection with man and machine in other types of military applications, the results of which may be applicable to the submarine.

In order to understand the objectives of the survey better and to become familiar with some of the physical and operational features of submarines, the panel members and the authors attended a meeting at the Submarine Base in New London, Connecticut on 8 and 9 December, 1947. All participants are greatly indebted to Rear Admiral P. F. Lee, Chief of Naval Research, ONR, and to Rear Admiral James Fife, Commander Submarine Force, Atlantic Fleet, who made this meeting possible. Likewise, grateful acknowledgement is made to the staff of ComSubLant, who participated in the indoctrination lectures and demonstrations. Special indebtedness is acknowledged to Captain Roy S. Benson and Captain Thomas L. Willmon, who so ably discussed operational and other problems, and to Dr. John Ide of the Underwater Sound Laboratory for discussing sonar and other types of detection equipment. Professor Walter S. Hunter, Brown University, and Dr. Eugene F. DuBois, Cornell University Medical School, representing the Committee on Undersea Warfare, graciously gave of their counsel at this meeting and provided many helpful suggestions.

During the course of the preparation of the chapters for this volume it was necessary for several of the authors to visit submarine bases and observe during trips on submarines. To the officers and men who so willingly gave of their time and cooperation to make these visits possible and profitable, the panel and authors are indeed grateful. The Panel is especially grateful to Commander William D. Groverman and his staff, Undersea Warfare Branch, Office of Naval Research, for their assistance in these and many other matters.

Although extremely busy with other duties, Dr. Francis W. Irwin, Editor of the Journal of Experimental Psychology, has shouldered the burden of the general editing of the manuscripts for publication in this volume. The indexes were prepared by Dr. James C. Diggory, University of Pennsyl-

vania. Each of the panel members assumed responsibility for securing the authors for certain sections of the report, and critically read the manuscripts prior to their submission to the general editor. However, primary credit and responsibility for the content of the various chapters must rest with their respective authors, who have the grateful appreciation of the Committee and the Panel for carrying out their difficult assignment so faithfully and well.

One of the major problems in surveying the literature relevant to submarine warfare was the matter of securing reports of research conducted during the war. There were problems of clearance for use of classified reports, but, perhaps more significantly, the sheer difficulty in finding and circulating these reports to the authors who needed them. In this regard acknowledgment should be made to the officers in the Office of Naval Research, and in the Bureau of Medicine and Surgery, who gave their cooperation and aid. In the handling of these matters Mr. Hugh Flynn, Technical Aide to the Committee on Undersea Warfare, was especially helpful.

Finally, tribute should be paid to Professor Walter S. Hunter and to Dr. Eugene F. DuBois, psychologist and physiologist, respectively, for their constant encouragement and valuable advice throughout this endeavor. The staff of the Office of the Committee on Undersea Warfare deserve special commendation for the efficiency with which they have handled the many details and arrangements necessary to the successful completion of this volume. In this connection, Mr. John S. Coleman, Executive Secretary of the Committee on Undersea Warfare, and Mr. Hugh Flynn, were of inestimable assistance with all phases of the preparation of this volume.

DONALD B. LINDSLEY, *Chairman*
Panel on Psychology and Physiology
Undersea Warfare Committee

FOREWORD

The Psychology and Physiology Panel of the Undersea Warfare Committee has as its mission the preparation of a general summary of the present status of knowledge with reference to the role of the human factor in undersea warfare. The present Report, although extensive, can only serve to outline the most essential aspects of the field for the general information and guidance of those whose decisions must shape the nature and extent of a research and development program in this field. The Report cannot do justice to all of the many aspects of human nature which must be considered in selecting and training men, in designing equipment for their operation, in safeguarding morale, or in waging psychological warfare. The Report will, however, serve a purpose of great value to the extent that it focuses the attention of military men on the advantages to be gained by an alertness to the existence of human-factor problems and by an insistence that active steps be taken for their solution. Although training in physics, chemistry, mathematics, and engineering is extensively given in military colleges and although many officers competent in these fields are available for materiel research and development, the opposite is true with respect to the science of human behavior. It is as though wars were fought by machines and not by men using machines, so that a knowledge only of the physical sciences need be required of responsible officers. At the present time there is a fairly wide-spread acceptance among military men of the importance of human-factor research for military operations, but there are almost no military men professionally competent to direct or conduct research in this area, a situation which contrasts markedly with the large numbers competent in the physical sciences and engineering. Perhaps the circulation of this Report will lead the Services to increase the number of officers assigned to postgraduate study in psychology to the end that the Services may, themselves, have personnel who are alert to the ways in which a science of human behavior can be utilized to increase efficiency and who are capable of directing research and development in this field. The National Military Establishment cannot expect satisfactory results in the scientific exploitation of the human factor if all but complete dependence for the formulation and solution of psychological problems continues to be placed on civilian scientists unaided by military officers themselves professionally competent in the field of human resources.

Undersea warfare in some possible future conflict will involve, not only the pro-submarine operation of equipment by highly selected and trained personnel, but also the operation of all types of anti-submarine equipment from the air, the surface, and below the surface. There is no indication that such operations can be conducted mechanically and by push-buttons without human operators. Rather, the high speeds of ships, planes, and projectiles, the prolonged submergence of submarines, in short the whole accelerating tempo of war, will place an extraordinary premium on the scientific utilization of the capacities and limitations of operating personnel. The present Report, although focused on the submarine problem, is not limited in scope to this one Service, and the principles which are presented are applicable in various ways throughout the National Military Establishment. This very fact should underline the importance of an increased coordination within the Services of the research and development effort in the field of the human factor. For example, research on cockpit design for planes has a relevance for the design of submarine control rooms. Systems research on the CIC for surface ships involves findings that should be considered on smaller ships and on submarines as well as vice versa. Problems of sleep, diet, habitability, and combat strain vary from one situation to another, but they are always present to affect efficiency. Dial designs that are desirable for the Field Artillery may well find an important place on shipboard equipment, and the limitations of the human factor which have led to the use of data transmitters and servo-mechanisms in anti-aircraft fire might well be considered by the Field Artillery. Even the principles of psychological warfare which can be developed and turned against the civilian and military personnel of an enemy can be used to develop countermeasures which will protect and enhance the morale of one's own group.

The utilization of the results derived from scientific studies of the human factor in World War II was frequently followed by an increase in personnel efficiency of from 15% to 35%. This was true in voice communication, in code training, in reading gun sights, and in the firing of flexible guns on the B-29. With a large military establishment, enormous gains or losses in total man-hours hinge upon correct decisions with reference to selection and training. Equipment designed without a nice regard for the capacities and limitations of the human operator becomes a long-continuing source of inefficient performance. This regard should be exercised during initial concept and preliminary design stages, if full benefit is to be realized. Too often the expert in human engineering is consulted at a time when only relatively minor changes in design or function are possible. On the other hand, in the relatively few cases where the psychologist has been made a member of the team of physical scientists, engineers, tacticians, and educators, rapid and significant progress has been made. In 1944 a distinguished Naval officer of engineering could write as follows: "As a result of such scientific applications of psychology in the selection and training of personnel, we have killed more submarines, have used radar better, have better gunners, better engineers, better communications both by voice and by code, have been shooting down more airplanes because our rangefinder operators were selected and trained better."

Should another war come, victory may well be, not on the side of the strongest battalions, not even on the side of the best guided missiles, but on the side which has gained a vital 10 percent in the successful handling of human factor problems.

WALTER S. HUNTER, *Member*
Undersea Warfare Committee

TABLE OF CONTENTS

PART V. HABITABILITY: A. PHYSIOLOGICAL FACTORS

PART VI. HABITABILITY: B. PSYCHOLOGICAL FACTORS

PART VII. EMOTIONAL STABILITY AND ADJUSTMENT

PART VIII. SELECTION AND TRAINING

PART IX. PERSONNEL RESOURCES

* Chapters A, B, and C, are bound separately in a classified supplement to this volume and may be made available to authorized persons upon application to the Chief of Naval Research, Navy Department, Washington, D. C.

PART I.

GENERAL VISUAL PROBLEMS

Chapter 1

HOW WE SEE: A SUMMARY OF BASIC PRINCIPLES

A. CHAPANIS

The Johns Hopkins University

General Introduction

The seven chapters which comprise this section on Visual Problems discuss experiments and sets of visual data which should assist the naval engineer, scientist and tactician in designing machines and operations to use the eyes more effectively.

The chapters which follow this one deal with reasonably specific applied problems. The present chapter serves both to orient the reader with respect to certain general problems of visual research and to provide a background of visual information to which ensuing chapters will refer and upon which they will build. In particular, this chapter deals briefly with units and methods of measurement in visual science, and then in more detail with a description of certain visual processes and functions which enter into all specific tasks from map or dial reading to visual search.

Although the scientific literature on vision is enormous, the engineer or applied visual scientist who attempts to put this information to practical use soon discovers that there are rather conspicuous gaps in our knowledge. He observes gaps not only in many areas of applied research but also in our basic research information about vision. Such data as he does find are often deficient because of two major faults which characterize many of our visual studies: they have been done to satisfy just one or two theoretical points, and they have used but few subjects. As regards the first point, the laboratory scientist appears particularly loath to repeat the work of his colleagues. Once the general nature of some visual function is known, his interest in that particular problem, as a problem for research, is over. The applied visual scientist therefore must frequently be content with data in which important parameters have been incompletely studied, or which can be put to use only by making questionable assumptions.

The second kind of deficiency is also serious, because the solution to a practical problem often requires information about the visual characteristics of the *average* person. In addition to questions of reliability which can be raised about experiments done on small samples of people, many visual studies are open to the criticism that they have been made on biased samples—on subjects purposely selected to have excellent visual acuity, no visual defects, etc. Samples of such individuals are hardly representative of people in general, or even of servicemen in general. This means that the repetition of many studies, although not as stimulating as the exploration of new problems, is definitely needed to provide the practical scientist with better normative data.

As the well-established, useful facts about vision are reviewed in this and the following chapters, numerous examples of research needs will be pointed out and discussed. It is hoped that they may stimulate investigations on what the writers believe are fruitful problems.

The Nature and Measurement of Visual Stimuli

Visible Radiant Energy

The radiant energy to which our eyes are sensitive is composed of electromagnetic radiations produced by electric charges moving through space at a very fast rate (roughly 186,000 miles per second). Although it is difficult to construct mechanical analogues of electromagnetic radiations, it is conventional and convenient, for our purposes, to talk about them as though they traveled in a wave form. One fundamental way of measuring and classifying radiant energy is in terms of the distance from pulse to pulse of the vibration, i.e., the wavelength of the radiation. These wavelengths cover an enormous range, from ten-trillionths of an inch (the cosmic rays) to many miles in length (Fig. 1). So far the observer may also say that the light is colored, or simply that he sees a color. It is very seldom that the ordinary person sees radiant energy composed of a single, or very few, wavelengths. When the eye is exposed to relatively homogeneous radiation, however, the observer usually reports seeing a color—one of roughly 150 identifiable colors in the spectrum. Visual scientists do not have names for all these different colors, but some of the more familiar ones are labelled in Fig. 1.

The wavelength of visible radiant energy is commonly measured in millimicrons (mμ), or Ångström units (Å). Although we

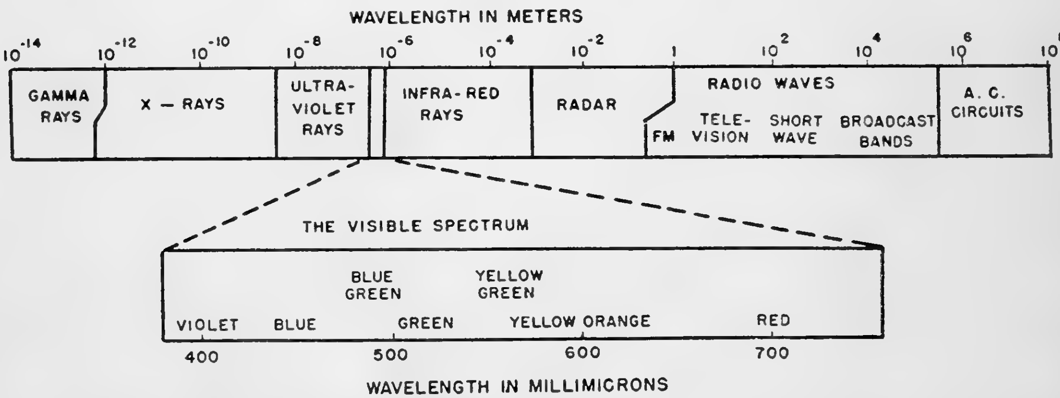

Fig. 1. The electromagnetic (radiant energy) and visible spectra. The various regions are defined somewhat arbitrarily because this is all the same kind of energy. Different observers will also disagree somewhat about the precise limits of the visible spectrum and the precise locations of the colors shown here.

as we can tell, all of these are the same kind of radiation physically. They differ only in wavelength.

Although radiant energy is very much the same physically, not all of it is visible. Somewhere in the middle of the spectrum, between 16 and 32 millionths of an inch in length, are those radiations our eyes respond to. The bundles of radiant energy which stimulate our eyes are usually composed of many wavelengths, and, when the eye is exposed to such radiation under certain viewing conditions, the observer usually says that he sees light. Depending on the distribution of radiant energy in the bundle, shall use only the former measure in this series of chapters, the following equation shows how these two measures are related to each other and to other more familiar units of length.

$$1 \text{ m}\mu = 10 \text{ Å} = 10^{-3}\mu = 10^{-9} \text{ meters}$$

Units and Problems in Photometric Measurement

Throughout the current set of chapters, a limited number of photometric units and concepts will be used. These are summarized and defined in Table I.

Photometric Concepts and Nomenclature. Unfortunately, the general subject of measurement in the visual sciences is in a state of flux at the present time. Some of the problems which are faced in this field are hinted at by the information given in Table I. In the first place, nomenclature has never been standardized to the satisfaction of all visual scientists—a situation which does not make for the easy reading of research reports. The most constructive step in the direction of standardization came with the recommendations made by the Committee on Colorimetry of the Optical Society of America in 1944 (**18**). This committee recommended the term illuminance to refer to the density of luminous flux falling on a surface, and luminance to refer to the luminous flux per unit solid angle emitted per unit projected area of a source (or reflected per unit projected area of a surface). The lag in the adoption of these terms is perhaps best evidenced by the fact that the authors of the ensuing visual chapters for the most part are still using the older nomenclature of illumination and brightness, and that they use units of luminous emittance when, strictly speaking, they should have used units of luminance. To sketch the details and background of this nomenclature problem would delay us unduly here, but these matters are considered to be of such consequence that they have been made the subject of a special appendix to this chapter. The reader is urged to consult it.

TABLE I

Photometric Units and Nomenclature Used in This Volume

Nomenclature		Units of Measurement		
Most common traditional term	Term recommended by O.S.A. Committee on Colorimetry in 1943 (18)	Name	Abbreviation	Definition or Equivalent
Illumination	Illuminance	Foot-candle	Ft.-C	Illuminance falling on the inner surface of a sphere with radius of 1 foot and a point source of 1 International Candle at its center
		Mile-candle		3.587×10^{-8} Ft.-C
Brightness	Luminous emittance	Foot-Lambert or equivalent foot-candle	Ft.-L	Total amount of light emitted in all directions from a perfectly diffusing, perfectly reflecting surface receiving 1 Ft.-C
		Millilambert	mL	0.929 Ft.-L
		Micro-micro-lambert	$\mu\mu$L	1×10^{-9} mL
Brightness	Luminance	Candle per square foot	C/ft.^{2}	Amount of light emitted in the direction of the eye from a perfectly diffusing, perfectly reflecting surface receiving π Ft.-C

Multiplicity of Photometric Units. A second problem suggested in Table I is the multiplicity of units which are being used to measure the same magnitudes. There are, for example, some 10 different units in reasonably common use for the measurement of luminance. This, too, is an unbearable situation and one which should be resolved by appropriate standardizing action by visual scientists. It should be noted that among the units used in the present chapters, the foot-lambert and the milli-lambert are essentially equivalent, and that the mile candle and the micro-microlambert are convenient for use in special studies because of their small size.

Photometry. The most commonly used technique in photometric measurement consists of comparing the light to be evaluated with some standard, controllable source, using the observer's eye as a null indicator. This technique is perfectly satisfactory for measurements at high and relatively high luminances, but at low luminances it has limitations, particularly when the luminous flux is restricted in its spectral composition. These limitations arise because of the properties of the particular visual cells used in making the photometric comparison. Problems of low luminance photometry will arise in a number of places throughout this and other chapters. They deserve investigation in and of themselves, for unless photometric techniques are standardized, research results at low luminance levels will always be difficult to compare. For a further and more systematic discussion of this stumbling-block to sound, low luminance visual studies, the reader is again referred to the appendix to this chapter.

Dark Adaptation and Night Vision

From the standpoint of its function, the eye may be considered as two separate systems. One system operates mainly at ordinary illuminances, such as those encountered throughout the day and in normally lighted rooms at night. The other system takes over at very dim illuminances, under starlight and moonlight conditions. There are some important differences in the way seeing is accomplished when each system is in operation and, during the last war, many articles and pamphlets were written on the subject of how to see at night. In the main, these publications were essentially correct—as far as they went. They erred chiefly in omitting many important experimental findings. These errors of omission were excusable at the time they were committed, but unfortunately they tend to be perpetuated and foster many incorrect, or only partially correct, ideas about dark adaptation and night vision. The following discussion, it is hoped, will rectify this situation to some extent.

Cone and Rod Vision

Histological studies have provided us with an anatomical explanation for the differences between day and night vision. The back part of the eyeball, the retina, contains two kinds of light-sensitive cells, the cones and rods. Several lines of experimental evidence show fairly conclusively that the cones are the receptors associated with day vision, and the rods those associated with night vision. There is also good reason to believe that individual cones and rods differ in sensitivity and that both types of cells operate over a wide range of illuminance. In general, however, vision in daylight is largely a function of the cones, while the rods are mainly responsible for vision at levels of illuminance below full moonlight.

Distribution of Cones and Rods in the Retina. Cones and rods are not distributed evenly throughout the retina. The classical data on this point come from the histological examination of a human retina by Østerberg (**72**) in 1935. Østerberg counted the number of cones and rods in areas of a standard size at 164 different positions on the retina. The data plotted in Fig. 2 are for those areas which were sampled on or close to a horizontal line through the center of the retina. Although these data are the result of a prodigious amount of labor, it should be understood that they were obtained on a single eye.

Fig. 2 shows that the density of cones is very much the same throughout most of the retina. In a small area around the fovea, however, the cones increase greatly in number and reach an estimated maximum density of 147,000 per square millimeter at the center. Rods, on the other hand, are entirely missing in the fovea (in Østerberg's specimen, in a region about 0.6° in diameter). Outward in all directions from the fovea, they increase in number and reach their greatest density in a region about 20° from the center. Beyond this region, the rod population decreases out to the extreme periphery of the eye. Very much the same pattern is obtained in other directions from the fovea, e.g., above and below it. It is also clear from Fig. 2 that the rods outnumber the cones throughout most of the retina. Østerberg estimates that the eye contains between 110,000,000 and 125,000,000 rods and between 6,300,000 and 6,800,000 cones.

Nervous Connections to the Cones and Rods. There are also some important differences in the kinds of nervous connections which link the cones and rods with the optic nerve. In the fovea, cones have their own individual nervous connections into the optic nerve. It is probable that many cones in the periphery also have their own individual lines of communication, although it is more common to find several cones, or several cones and rods, linked together on a single nerve pathway in the outer regions of the retina. Rods, however, are never found to have individual nervous connections; they are always linked together in groups to single nerve fibers.

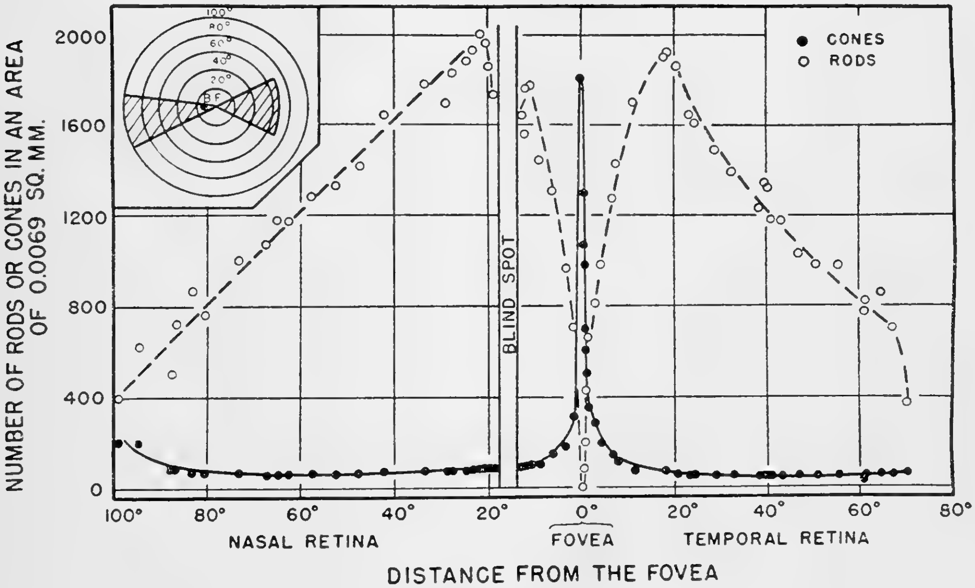

Fig. 2. The density of cones and rods on or near the horizontal meridian through a human retina. The inset is a schematic map of the retina showing *F*, the fovea, and *B*, the blind spot. The striped area represents the regions of the retina which were sampled in obtaining the counts plotted here. (Data from Østerberg, **72**)

The relative concentration of the cones and rods in different parts of the retina, and the kinds of nervous connections these cells have with the optic nerve, determine in part certain important visual characteristics. As will be pointed out later (Fig. 17), visual acuity is best when the image of an object falls directly on the fovea. It is here, in the fovea, that we find the greatest density of cones. In addition, the pin-point, one-to-one, transmission of nerve impulses from individual foveal cones into the optic nerve undoubtedly helps to maintain the accuracy of detail originating there. Visual acuity in the periphery of the eye is much less acute than in the fovea due, in part at least, to the fewer cones and the coarser grouping of them into individual nerve channels.

Sensitivity of the Retina to Light. As might be anticipated on the basis of the foregoing information, the sensitivity of different parts of the retina to radiant energy also varies considerably. The results of a very careful study on this function by Wentworth (**96**) are shown in Fig. 3. The test lights used in the experiment were 1° 16′ in size, and were isolated from a spectrum at the following wavelengths: red (R) 672.5 mμ; yellow (Y) 581.5 mμ; green (G) 522 mμ; and blue (B) 468 mμ. Two obvious conclusions which can be drawn from these data are (1) that the eye is most sensitive to radiant energy at some distance from the fovea; and (2) that the eye is not

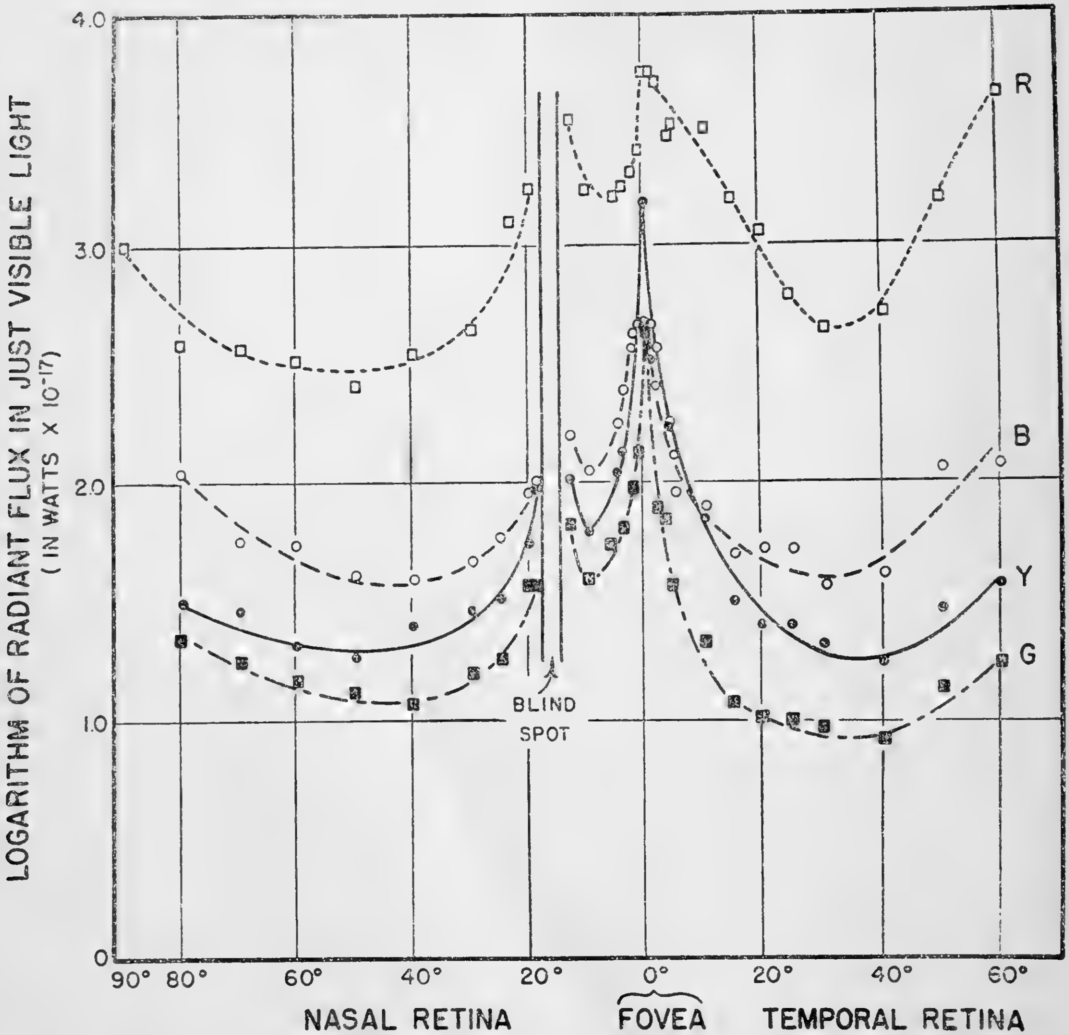

Fig. 3. The least amount of radiant flux perceptible to the human eye at various retinal locations. The *R*, *B*, *Y*, and *G* are wavelengths of 672.5, 468, 581.5, and 522 mμ, respectively. (Data from Wentworth, **96**)

equally sensitive to radiant energy from different parts of the spectrum.

Although the data cited above are very detailed and informative, they were, unfortunately, obtained on only one subject. Similar data by Sloan (**84**) on 101 subjects, ranging in age from 14 to 70 years, are shown in Fig. 4. These measurements were made with a 1°, white test light. The solid line in Fig. 4 represents the average threshold that Wentworth's eye was atypical may account for the discrepancies. In any case, one thing is clear: the dark-adapted eye is most sensitive at some considerable distance from the fovea. It is interesting to note, incidentally, that these sensitivity curves parallel fairly closely the rod density curves shown in Fig. 2. The eye appears to be most sensitive in those regions where the rods are densest.

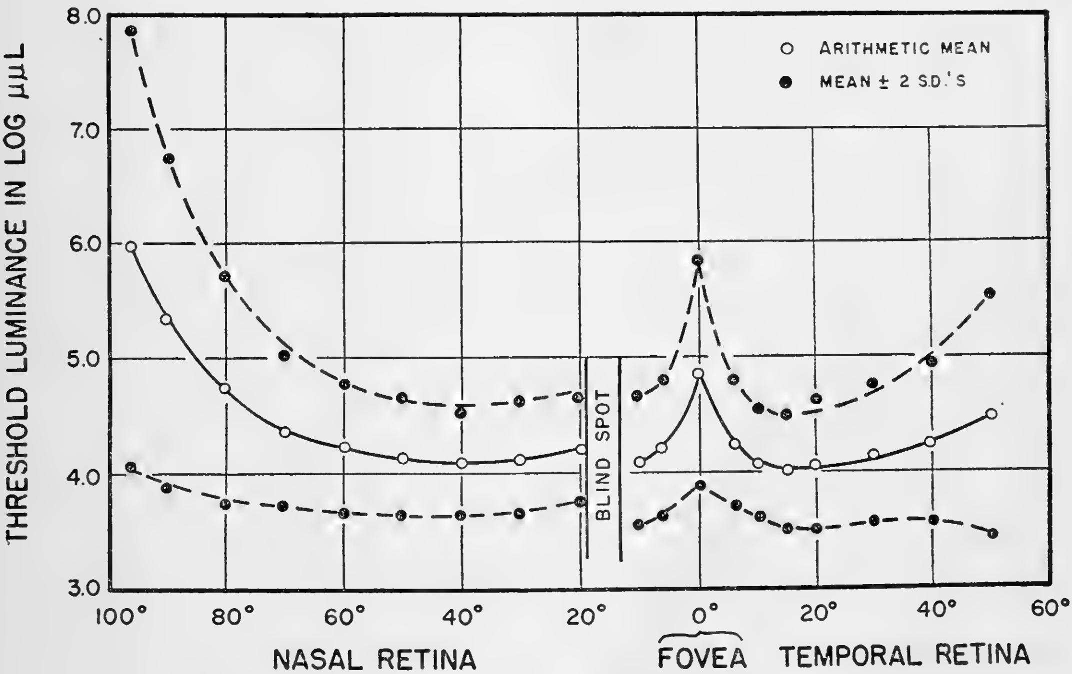

Fig. 4. Luminance of the just perceptible light at various retinal locations. The solid line is the average curve for 101 subjects; the dotted lines enclose the measurements of 95 percent of the subjects. (Data from Sloan, **84**)

values for the 101 subjects. The dotted lines are the average values plus and minus two standard deviations; they thus enclose the measurements of 95 percent of the subjects tested. This study shows the most sensitive part of the retina to be somewhat closer to the fovea than is indicated by Wentworth's data. It is not clear why these two sets of data do not agree more closely. Differences in the colors of the test lights used, differences in the units of measurement employed, or the possibility

Sensitivity of the Cones and Rods to Radiant Energy. The data by Wentworth cited above show very clearly that the eye is not equally sensitive to radiant energy from different parts of the spectrum. This characteristic of the eye has now been thoroughly studied and is usually quantified in the form of photopic and scotopic relative luminosity curves as illustrated in Fig. 5. The scotopic relative luminosity curve in Fig. 5 is an average curve for 48 subjects studied by Hecht and Williams (**39**).

Each subject was required to vary the intensity of the light produced by radiation of each of the wavelengths shown (412 mμ, 455 mμ, 486 mμ, etc.) until it matched a very dim white light very close to the dark-adapted threshold of the eye. The curve is called a relative luminosity curve because it shows the relative luminosity (or amount of light) produced by the various wavelengths in an equal-energy spectrum. Put in other words: if the completely dark-adapted eye were to look at a very dim, equal-energy spectrum, the different parts of the spectrum would not appear to be equally luminous. The most luminous part of the spectrum occurs at 511 mμ; 462 and 555 mμ appear half as luminous as 511 mμ, etc.

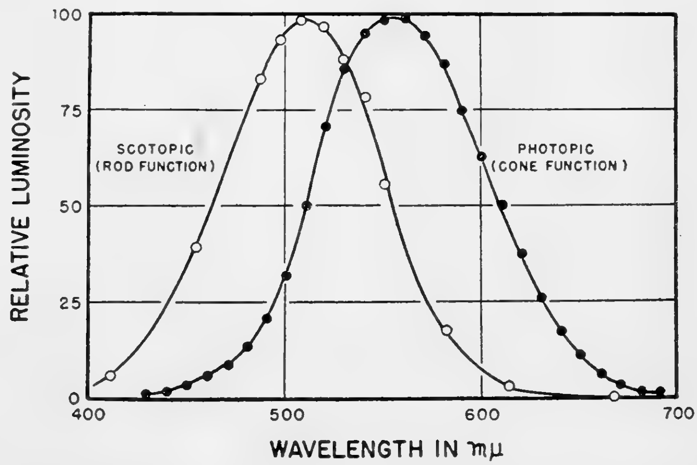

Fig. 5. Photopic and scotopic relative luminosity curves. The photopic data are from Gibson and Tyndall (**28**); the scotopic data from Hecht and Williams (**39**).

The photopic relative luminosity curve shown in Fig. 5 represents the average data for 52 observers studied by Gibson and Tyndall (**28**). Although their experiment differs somewhat from the one of Hecht and Williams in its details, the resulting curve has essentially the same meaning with this exception: the photopic relative luminosity curve shows the relative luminosity of the various wavelengths in an equal-energy spectrum when the intensity of the spectrum is well above the cone threshold. The scotopic curve is thus an expression of rod sensitivity, the photopic curve an expression of cone sensitivity.

The Purkinje Shift. Although photopic and scotopic relative luminosity curves appear in almost all secondary literature on vision, it is not generally understood that these curves show the sensitivity of the eye under extreme conditions. The photopic curve is found when the eye is light-adapted to a fairly high luminosity level; the scotopic curve results when the eye is completely dark-adapted. When the luminance is decreased from photopic to scotopic levels, as for example from afternoon illuminancse through twilight to dusk, the transition from cone to rod vision is not abrupt. The eye does not start with photopic sensitivity and switch suddenly to scotopic sensitivity. Furthermore, the transition is gradual even if the change in illuminance is abrupt. Fig. 6 shows the results obtained by Walters and Wright (**95**) on one of their two subjects. The different curves were obtained at five selected luminance levels ranging from cone levels to rod levels. Measurements were made on a retinal area 10° from the fovea. Notice the *gradual* shift in the luminosity curves as the luminance changes from the cone to the rod levels. Notice also that there is an increase in the relative luminosity of blue and blue-green lights (wavelengths below approximately 510 mμ) as adaptation proceeds from cone to rod levels. This shift in the relative luminosity curve is variously called the Purkinje phenomenon, Purkinje effect, or Purkinje shift, for the famous Bohemian physiologist who discovered it.

are used to provide cone levels of illuminance on a working area, the eye of an observer nearby will have sensitivity values as illustrated by the photopic relative luminosity curve. As the observer moves away from the light sources, the illuminance falling on his eye will decrease according to the inverse square law, and the sensitivity of his eye will gradually undergo the

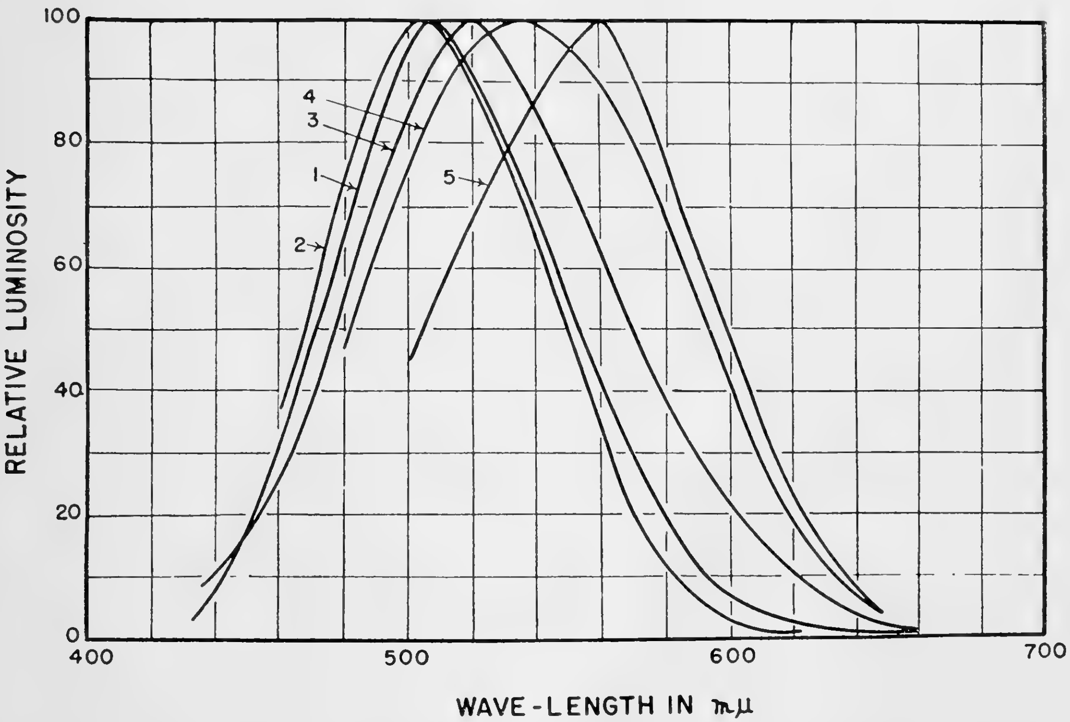

Fig. 6. Relative luminosity curves at five luminance levels (1 is the lowest luminance; 5 the highest). The gradual shift in the luminosity curve is called the Purkinje shift. (Data from Walters and Wright, 95).

Practical Import of the Purkinje Shift. The practical significance of the Purkinje shift is apparent. If variously colored lights, all of equal size and intensity, are used to provide cone levels of illuminance at night, these lights will not appear to be equally luminous to a distant observer. Green, blue-green and blue lights "outlive" other colors in visibility as the distance between the light source and observer increases. The reason is now easy to understand. If the variously colored lights

Purkinje shift, i.e., will become relatively more sensitive to the shorter wavelengths. Thus, although the lights may all have appeared equally intense at short distances, the green, blue-green, and blue ones will be visible for much greater distances. This, however, is true only if the distances involved are much greater than the largest dimension of the light source.

The Purkinje Shift in Different Retinal Locations. Since there are no rods in the fovea, it might be anticipated that the

Purkinje shift would not occur there. This is essentially correct. Walters and Wright have shown that relative luminosity curves measured in the fovea change slightly as the luminance is changed from full daylight to full twilight conditions. The change is very small, however, as compared with the full range of the Purkinje shift which occurs in the peripheral regions of the eye. Because of the varying density of the rods in different retinal locations, (see Fig. 2), it might also be anticipated that the Purkinje shift would occur at different rates in different parts of the eye. This expectation is also in accord with the experimental facts. Walters and Wright studied the retina 3° and 10° from the fovea. The magnitude of the Purkinje shift at 3° was not quite as great, and did not proceed at as rapid a pace, as at 10°.

Absolute Sensitivity of the Cones and Rods. The photopic and scotopic relative luminosity curves shown in Fig. 5 are frequently misinterpreted in another important respect: it is not generally appreciated that the curves shown there are *relative* curves; each is drawn relative to its own maximum. This misunderstanding leads to the frequent mis-statement that the rods are not as sensitive to red as are the cones. Several lines of evidence, in particular recent experiments by Chapanis (**7**, **8**), and Wald (**91**), show conclusively that the rods are, in fact, just as sensitive to red as are the

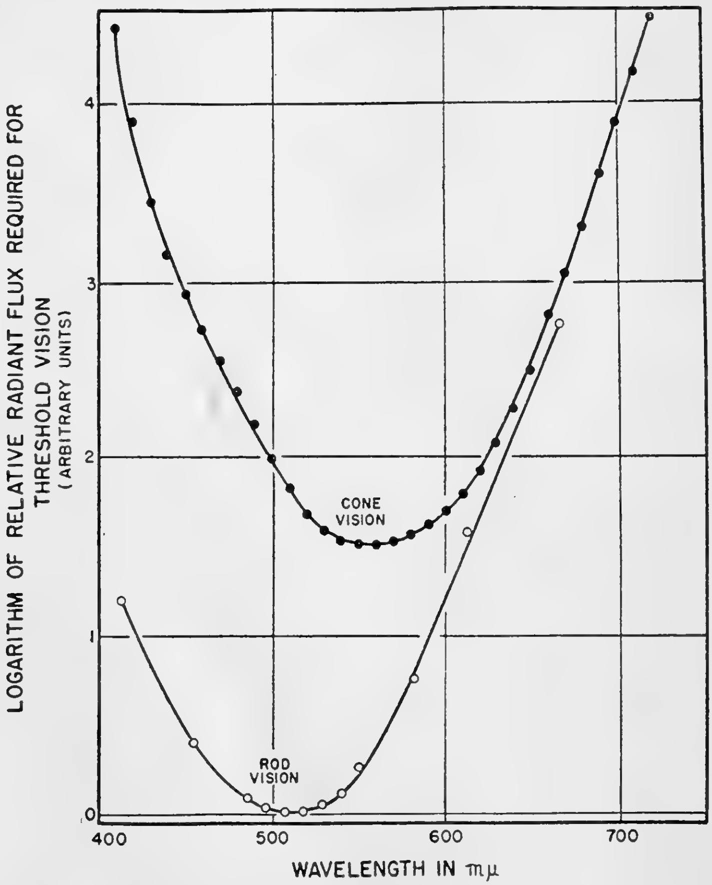

Fig. 7. Relative amounts of radiant flux required to stimulate the rods and cones. These curves are replotted from the data shown in Fig. 5.

cones. If the relative luminosity curves are replotted in terms of the amounts of energy involved, they assume the shapes shown in Fig. 7. Although the curves in Fig. 7 were actually computed from the data of Fig. 5, Wald's data (**91**) show that this formulation is in close agreement with the experimental facts. This kind of a plot illustrates three points clearly: (a) in progressing from cone vision to rod vision, there is a shift in the region of maximum sensitivity (from 555 mμ to 511 mμ)—the Purkinje shift, (b) throughout most of the spectrum the rods require much less radiant energy for vision than do the cones, and (c) the rods are about as sensitive as the cones to radiation from the long wavelength (red) end of the spectrum.

Application to Low Luminance Photometry. Having reviewed very briefly the behavior of the eye at different luminance levels, we are now in a better position to appreciate why the problem of low luminance photometry is such a difficult one. The Purkinje shift is at the root of the trouble, and the situation is complicated by the fact that the Purkinje shift is different for different retinal positions. The development of standard techniques and procedures to meet these difficulties will require considerable work.

Color Vision. The final point of interest with regard to cone and rod function is that color vision is possible only with cone vision. When only rods are operating, it is possible to distinguish between light and dark colors only in terms of the intensity of the reflected or transmitted light. In rod vision, then, all colors appear as a series of lighter or darker grays. This does not mean, of course, that color vision is impossible at night. Common experience in observing neon signs shows that colors can be seen at night; but this is true only if they are above cone thresholds.

Dark Adaptation

Going suddenly from a brightly lit environment into darkness is a common experience. Moviegoers endure it almost every time they attend the theater, and are frequently embarrassed and annoyed by the attendant consequences. Immediately after being plunged into darkness, it is very difficult to see things. Gradually, however, more and more details of the environment become perceptible, and, after a period of from 10 to 30 minutes, very dim objects can be seen. This increase in the sensitivity of the eye is referred to as dark adaptation. In this sense, it refers to a process—the progressive increase in sensitivity of the eye when the external illuminance has been decreased. The term has also been used, however, to refer to the constant state of sensitivity reached by the eye when it has remained in total darkness for some time.

Instantaneous Threshold of the Eye After Light Adaptation. Contrary to the impression one gets from reading war-time literature on dark adaptation, the eye is not completely insensitive to all light immediately after it has been plunged into darkness. It can see objects which are considerably dimmer than the general level of luminances to which it had been adapted. If this were not the case, of course, the eye would not be able to see into deep shadows. Instantaneous threshold measurements have been made by Blanchard (**3**), and a portion of his data are shown in Fig. 8. The experiment was done by adapting the eye to a large field illuminated to various levels. At a signal, the adapting light was turned off and the subject reported whether he could see a much dimmer light which replaced the adapting light. By many repeated trials of this sort, it was possible to find an instantaneous threshold luminance. Five different colors of adapting and test lights were used: white, blue, green, yellow and red. In any single trial, the adapting and test lights were always of the same color.

Notice that there are two fairly distinct portions to the three curves shown in Fig. 8. The data for the different colored lights are very much the same for adapting luminances above 0.1 millilambert (log = −1), i.e., above the cone threshold. At luminance levels below 0.1 mL (millilambert), however, the curves diverge. The yellow and green curves are not shown here, to keep from cluttering up the chart, but the yellow curve falls between the red and white curves and the green between the white and blue ones. Without attempting any detailed theoretical explanation of these data, it is easy to see that the Purkinje shift is at work here again.

For our purposes, it is important to notice that when the eye has been adapted to high luminance levels, e.g., 1000 mL, it can see luminances about one one-thousandth as intense immediately after it is plunged into darkness. When it has been adapted to a luminance of 1 mL (log = 0), it can see a light one one-hundredth as intense upon being plunged into darkness. The one difficulty with this set of data becomes immediately apparent at the lowest luminance values. If we were to trust the data implicitly, it would seem that at very dim adapting levels, the eye can see objects only if they are more intense than the light the eye had been adapted to. Although these very data have been used by Hecht (**35**) in support of his photochemical theory of vision, this inconsistency appears to have escaped him.

Further Research Needed. These data provide an excellent illustration of the difficulty produced by the lack of a consistent system of photometry for luminances below the cone level. The problem worried Blanchard, as is evident from his discussion. He says (**3**, pp. 85–86):

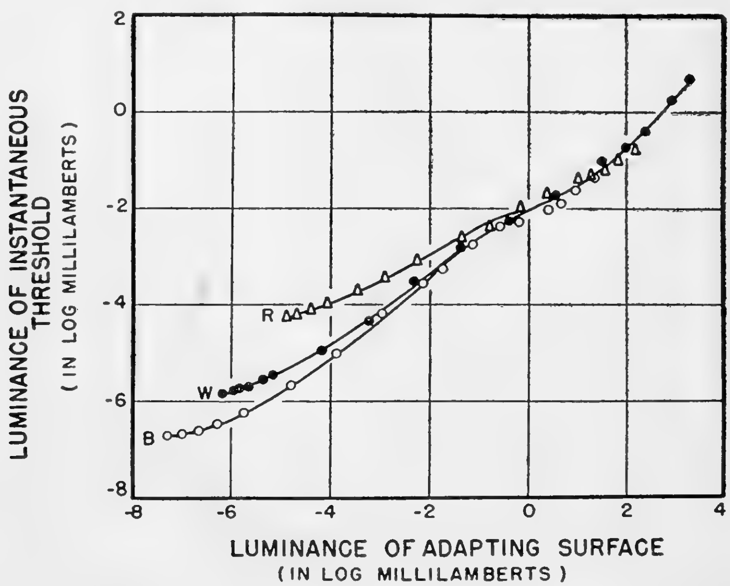

Fig. 8. The luminance of the just visible light immediately after lights of various luminance levels are turned off. (Data from Blanchard, **3**)

> In order to express the results consistently in the same unit of brightness it is necessary to take into account the Purkinje phenomenon. If two fields of different color are illuminated to the same apparent brightness and both cut down by equal amounts, the brightness will not decrease in the same ratio. For example, red will grow darker much faster than blue. But at very low intensities it is impossible to measure brightness by any photometric means, and without having a definite measure of the Purkinje effect for the different colors, the only feasible way of expressing relative intensities is in fractions of a certain measured intensity above the brightness at which the effect sets

in. In these experiments all the colors were measured photometrically at a brightness of 10 millilamberts which is safely above the Purkinje effect, and the lower intensities calculated from a knowledge of the filter densities.

Unfortunately, these data cannot be put to practical use because it is impossible to reconvert to any meaningful system of units from the data given by Blanchard. The same is true for many other studies of visual functions for different colors.

In addition to the specific problem of measurement which vitiates Blanchard's data, there are some others arising from his experimental conditions. Only a few of these will be mentioned here. Blanchard used a very large adapting field, about 70° in angular size, and a test field about 5° in size. This size of test field, unfortunately, stimulates an area somewhat greater than the fovea, i.e., a region of very uneven sensitivity (see Figs. 2, 3 and 4). Blanchard himself recognized this objection to his own data, but did not offer a solution to the difficulty. Still another difficulty with Blanchard's experiment is that it was apparently done with only one subject. Blanchard, as a matter of fact, does not say explicitly how many subjects were used, but the implication is that only one was used. Further, to be of greatest practical usefulness, this experiment should be extended to measure the instantaneous threshold with differently colored test lights, but with the adapting field always white. As we shall see later, the color of the adapting field has a very important effect on the subsequent dark adaptation of the eye and it is reasonable to infer that the instantaneous threshold would be similarly affected. Finally, Blanchard does not state explicitly how complete light adaptation was. This also is a very important point in experimental procedure, because, as we shall also see later, the subsequent sensitivity of the eye is markedly affected by the amount of previous light adaptation.

The above discussion indicates clearly that this is a problem which needs reinvestigation. Good basic data on this function would enable the engineer to solve several sorts of practical seeing problems. During the war, for example, radar screens were placed in fast day bombers. A very practical problem at the time was this one: how luminous must a radar screen be so that the operator can see it immediately after scanning the sky? Blanchard's data provide us with a reasonable basis for guessing, but they are hardly adequate for precise work.

The Typical Curve of Dark Adaptation. The process of dark adaptation is ordinarily traced by determining the dimmest light a subject can see at various times after the lights are turned out in a dark room. The data in Fig. 9 were obtained by Sloan (**84**) and are typical for this function. The test light in her experiment was white, 1° in diameter, and was situated in the nasal field of view, 15° from a fixation point. The solid line in Fig. 9 represents the average values for 101 subjects; the dotted lines include the thresholds for 95 percent of the subjects.

It is evident from Fig. 9 that there are two segments to the dark adaptation curve: an initial very rapid decrease in threshold which levels off at about 10 minutes, and another more gradual decrease in threshold which starts at about 10 minutes and continues for some time. The decrease in threshold which occurs during the first 8 or 9 minutes is due to an increase in the sensitivity of the cones and, to a small extent, to dilation of the pupil which increases the light-gathering power of the eye. The leveling off at about 10 minutes is now known to represent the cone threshold. If colored test lights are used to measure dark adaptation, the colors can usually be recognized up to this point (see Fig. 10). The decrease in threshold which occurs after 10 minutes is due to the increase in sensitivity of the rods. If colored lights are used, the colors cannot be distinguished at threshold after this time (see Fig. 10). It is from measurements such as these, incidentally, that rod and cone threshold data similar to those in Fig. 7 can be obtained.

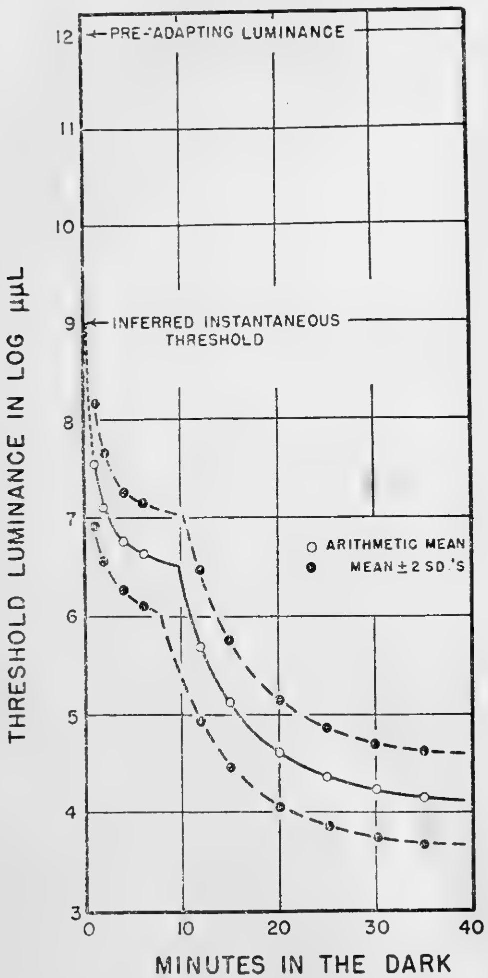

Fig. 9. The course of dark adaptation measured for 101 subjects. The solid line shows the luminance the average subject could see at various times after the lights were turned out. The dotted lines enclose the measurements for 95 percent of the subjects. (Data from Sloan, **84**)

The instantaneous threshold exhibited in Fig. 9 is an estimate from Blanchard's data. Notice that the dark adaptation curve itself extends over five log steps. This illustrates the enormous range of adaptation of the eye since five log steps correspond to luminances in the ratio of 100,000 to 1. If the full range of sensitivity of the eye is measured—from the pre-adapting luminance to the final rod threshold—the range is about eight log steps, or luminances in the ratio of 100,000,000 to 1!

An Inconsistency in Measurement. At this point we may indicate an inconsistency in concepts which is implied in the measurement of dark adaptation and, in fact, in all the data of this section whenever thresholds are reported (Fig. 4, for example). Light has been defined as "radiant energy . . . evaluated according to its capacity to produce visual sensation." But a stimulus below threshold, i.e., a stimulus which cannot be seen, does not arouse a visual sensation and so cannot properly be called a light. Yet most measurements of dark adaptation express the intensity of the stimulus in luminance measurements and show the scale extending below the visual threshold. The absolute sensitivity of the human eye must define the zero point of a luminance scale. For most practical purposes, this is a minor difficulty, but it is further evidence of the need for a consistent system of low-luminance photometry.

How Long Does it Take for Complete Dark Adaptation? A matter of some practical importance to the armed services is the length of time an individual must remain in darkness to reach maximum visual sensitivity. During the last war, this problem was important enough in certain types of military jobs, e.g., night lookout duty and night flying, for instructions to be issued requiring the men concerned to adapt for prescribed periods of time before reporting for duty. Many night-vision manuals written during this period stated categorically that 30 minutes in the dark were required to reach maximum visual sensitivity. This statement requires considerable amendment. In the first place, no experimental study to date has followed dark adaptation to completion. Semeonoff (**80**), however, studied the visual sensitivity of a subject who remained in darkness for six hours. He found that the curve of sensitivity changed, although very slowly, throughout the six-hour period. Any statement about the time required for adaptation, therefore, must use some arbitrary criterion of "complete adaptation." Perhaps some definition such as this would suffice: for all practical purposes, the eye is dark-adapted when the mean threshold value does not change by more than 0.1 log units in two consecutive five-minute intervals. As far as the writer is aware, no such standard criterion has ever been used in dark adaptation studies.

Dark Adaptation as a Function of Pre-Adaptation. The second reason we must be cautious about stating the length of time required for complete dark adaptation is that the course of dark adaptation is dependent on the amount of previous light adaptation. The data in Fig. 9 were obtained by having the subjects fixate a large surface with a luminance of 1,100 mL for three minutes before dark adaptation was started. This exposure to light before dark adaptation begins is called light adaptation, or pre-adaptation. A pre-adaptation period is used routinely in laboratory studies of dark adaptation to insure that subjects start with a constant amount of photochemicals bleached in the retinae of their eyes. It is a control used by the laboratory scientist to obtain more uniform and comparable dark adaptation records from different individuals. Since one subject may report to the laboratory after having been out in sunlight (with most of the photochemicals in his eye bleached by the strong light), while another may come in after an hour's work in a dim laboratory (with a relatively small amount of photochemicals bleached), the easiest way to start all subjects at the same level is to bleach their retinae the same amount by exposing them all to a very intense light. As a general rule, then, the pre-adapting lights used in laboratory experiments are very intense. The luminance of the pre-adapting light used by Sloan in her study is plotted in Fig. 9 to show how high it is.

As a practical matter, however, men who are required to work at night are very seldom exposed to luminances as high as those used for pre-adaptation in the laboratory. Pilots sitting around a ready-room

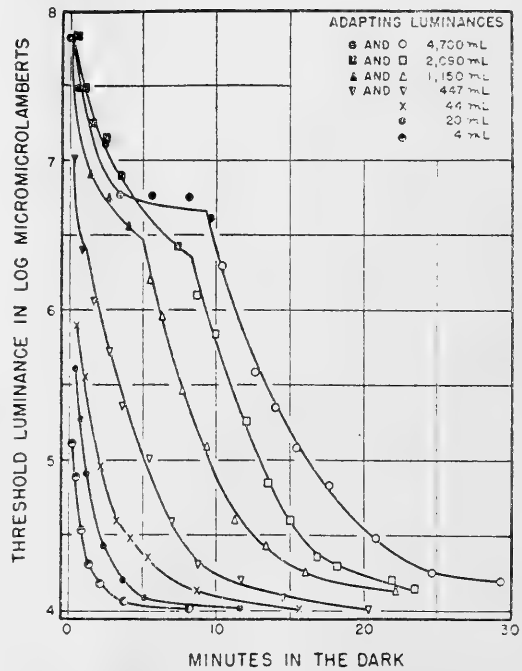

Fig. 10. Dark adaptation curves for one subject following exposures to lights of various luminances for four minutes. The solid symbols indicate that the color of the test light (violet) could be identified at threshold; open symbols indicate that the color of the test light could not be recognized. The test light appeared colorless at threshold following pre-adaptation to the three lowest luminances. (Data from Haig, **31**)

at night, for example, would probably be adapted to luminances of 1 to 30 mL *even if the ready-room were illuminated with several 50-watt bulbs.* The point is frequently overlooked that the time required for dark adaptation decreases rapidly as the luminance of the pre-adapting light decreases.

Data on this point come from a study by Haig (**31**). This investigator studied the course of dark adaptation following exposures to seven different luminance levels—4,700, 2,090, 1,150, 447, 44, 20 and 4 mL—for four minutes. He used two subjects, but one of them had a pathological eye condition and so cannot be considered typical. The data for the normal subject, Fig. 10, show how much more rapid dark adaptation is when the pre-adapting light has been dim. This is illustrated even better by the data in Fig. 11, which shows the time required for the dark adaptation curve to reach a value roughly twice the

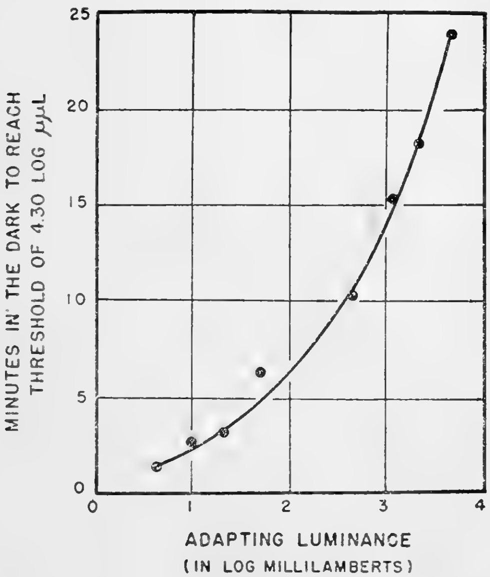

Fig. 11. The data of Fig. 10 have been replotted here to show the time required for the dark adaptation curve to reach a value near threshold as a function of the pre-adapting luminance.

threshold, (i.e., 4.30 $\mu\mu$L) as a function of the pre-adapting intensity. It is evident from Fig. 11 that near-threshold sensitivity is reached in a matter of a few minutes even if the pre-adapting light has been as high as 100 mL (log = 2).

It is important to note, incidentally, that the rod sections of the dark adaptation curves in Fig. 10 do not have the same form. When the luminance of the pre-adapting light has been dim, e.g., 4 to 20 mL, the dark adaptation curve drops much more rapidly than it does following exposure to higher values, e.g., 2,090 and 4,700 mL. We shall have something more to say about this difference in speed of adaptation in just a moment.

Dark Adaptation as a Function of the Duration of Pre-Adaptation. Another important determinant of the time required for dark adaptation is the length of the pre-exposure period. Data on this point, also from the study by Haig (**31**), are shown in Fig. 12. Only the rod portions of the dark adaptation curves have been plotted. In this part of his experiment, Haig used a

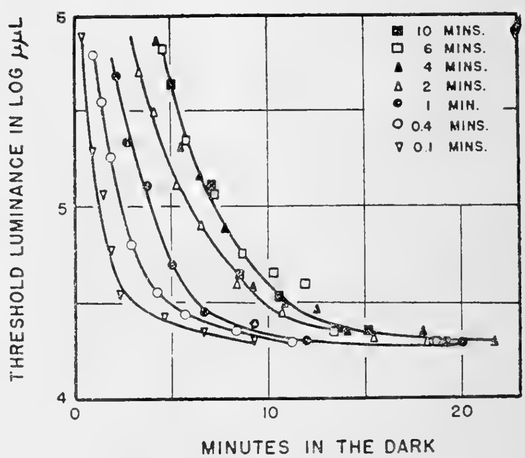

Fig. 12. Dark adaptation curves for one subject following exposure to light of 447 mL for various durations. Only the rod portions of the curves are shown here. (Data from Haig, **31**)

single luminance, 447 mL, but varied the duration of the pre-adapting period from 0.1 to 10 minutes. These data have also been replotted to show how long adaptation must continue to reach an arbitrary value close to threshold (Fig. 13). Although the data in Figs. 10 and 12 are for the same subject, they were obtained on different days, and the subject's sensitivity had evidently changed somewhat in the meantime. For this reason, it has been necessary to use a different criterion value in Fig. 13 from that used in Fig. 11. Nonetheless, the story is clear. With fairly short durations of pre-adaptation, dark adaptation is virtually complete in a very few minutes.

Different Rates of Dark Adaptation. It is also clear from Fig. 12 that the rate of dark adaptation following a very short pre-exposure period is more rapid than the rate following long pre-exposures. To get better evidence for this, Haig calculated a crude measure of the slope of each dark adaptation curve. He defined the slope as the reciprocal of the time required for the curve to drop 0.4 log units and computed it for two contiguous 0.4 log unit intervals, from 1.19 to 0.79 and from 0.79 to 0.39 log units above the threshold. All that is really important here, however, is that the slope tells us how fast the dark adaptation curve dropped. The higher the slope, the more rapidly the curve dropped; the lower the slope, the more slowly it dropped.

Slope values, plotted as functions of the luminance of the pre-adapting light and duration of pre-exposure, are shown in Fig. 14. This figure shows very clearly that dark adaptation may progress at different rates. It is much faster (a) when the pre-adapting intensity has been 44 mL or less, or (b) when the pre-exposure has lasted less than one minute. If the pre-exposure intensity has been higher than 44 mL or has lasted longer than one minute, dark adaptation proceeds at much slower rates. Haig (**31**) and Wald and Clark (**92**) have provided an explanation for this phenomenon in terms of two different kinds of photochemical reactions known to occur in the retina. We need not be concerned here with the details of these photochemical processes and the point is mentioned only to show that the entire story makes sense.

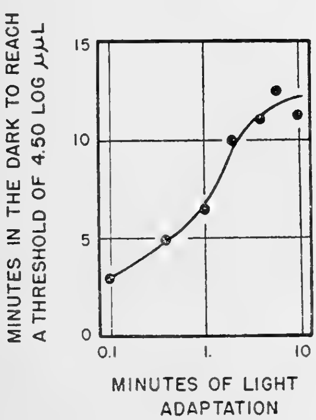

Fig. 13. The data of Fig. 12 have been replotted here to show the time required for dark adaptation to reach a value near threshold as a function of the duration of pre-adaptation.

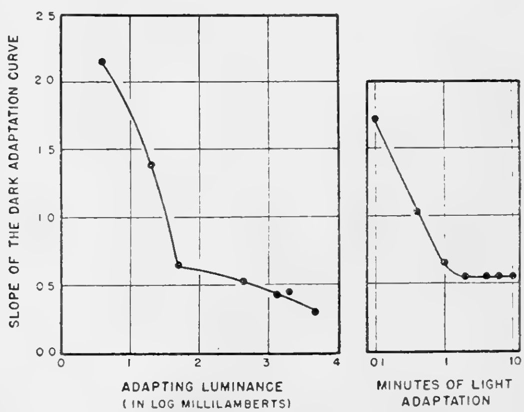

Fig. 14. The slope of the dark adaptation curve as a function of pre-adaptation luminance (left) and duration of pre-adaptation (right). These curves were computed from the data shown in Figs. 10 and 12. See the text for a fuller explanation.

The importance of this finding should be obvious. Whenever possible, the specifications for illuminance in ready-rooms should be such that they will permit the eye to dark-adapt at fast rates. Another implication, not so obvious perhaps, is that exposure of the eye to a dim light for a short period of time does not completely destroy all dark adaptation. It is important to remember that Haig's data were obtained with a luminance of 447 mL, which is a very high value. Fragmentary data from a study by Peckham (**73**) show that if the eye has been dark-adapted and is exposed to a luminance of 7.1 mL for eight seconds it returns to its fully dark-adapted condition within a minute. This means that no really great harm is done if a soldier or sailor has to turn on a very weak light for short periods of time to examine a map or message.

Further Research Needed. Although the

data of Haig are supported by the findings of a number of other studies (**36, 70, 98**), they are all fragmentary in one respect or another. Haig's study, as we noted, was done on one normal and one pathological subject. In his study of the effect of the pre-adapting intensity, he used only one exposure time, four minutes, which, according to his own data (Fig. 11), is not sufficient for full light adaptation. In his study of the effect of exposure time, he used only one adapting value, 447 mL. Virtually the same kinds of objections can be raised about every other study on this problem. As a result, it is impossible to obtain the critical values needed for any practical specification of ready-room illumination.

Further research on this problem would be of immense practical value. In particular, the following points should be observed:

(a) A greater number of subjects should be used to obtain better normative data.

(b) The effect of pre-adapting luminance should be studied with relatively long durations of pre-adaptation to insure complete light adaptation. Wald and Clark (**92**) and Müller (**70**) find that no further change in dark adaptation is evident if pre-exposures exceed 10 minutes. Here again, however, their studies were done with single luminance levels. Whether this is true for other luminance levels constitutes another problem in itself. If the subjects were completely light-adapted, the situation would be more comparable to that found in service conditions where men may sit around a ready-room or ward room for long periods of time.

(c) The effect of duration of exposure should be reinvestigated with a greater range of luminances.

(d) Haig believes that, within certain limits, the luminance of the pre-exposure light and the duration of the pre-exposure may be interchangeable. If this is so—and it seems entirely reasonable on the basis of other visual functions in which intensity and time may be interchanged—it would provide a single measure of "amount of light adaptation" which could be specified as a function of intensity and time. Amount of light adaptation, in turn, could then be related to the length of time required for subsequent dark adaptation. It should be evident that this conjecture is well-supported by our present data, but that these data are far too few to be put to any practical use.

With data of such investigations at hand, visual scientists would be able to specify with some certainty the length of time men must dark-adapt after having been in ready-rooms with known amount of illuminance. Present data suggest that this period will be of the order of a few minutes. The proposed new studies would also show how serious it would be for dark-adapted men to turn on dim lights from time to time.

Independence of Adaptation in the Two Eyes. It is now known that dark-adaptation proceeds independently in each eye. This fact is of considerable importance to the soldier or sailor because it means that if he must turn on a light for some reason he can close one eye to preserve the dark adaptation in that eye. When he turns the light off, the dark adaptation of the eye that had been closed is virtually undisturbed.

Color of Pre-Adapting Light. Still another important determinant of the rate of dark adaptation is the color of the pre-adapting light. If pre-adaptation is carried out with red light, dark adaptation proceeds very rapidly after the pre-adapting light is turned off. If, on the other hand, pre-adaptation is done with white, green or blue light, dark adaptation proceeds much more slowly. Although the validity of this relationship was challenged by Lowry (**55**) during the war, careful studies by Hecht and Hsia (**37**) and Rowland and Sloan (**79**) showed Lowry to be wrong. We may, therefore, accept it as verified fact.

The Use of Red Goggles. Red goggles were widely used during the last war to dark-adapt the eyes. Their use was so common that our discussion of this problem may be very brief. The red goggles serve two functions: (a) They reduce the overall amount of light reaching the eye so that the eye is light-adapted to a lower luminance level. (b) They admit only red light to the eye, with the result that subsequent dark adaptation is more rapid.

During the war, a great amount of research time was spent on the evaluation of different red filters in terms of their efficiency for dark-adapting the eye. In a general kind of way, it was appreciated that a filter was better for dark adaptation if it transmitted fewer of the shorter wavelengths. As a rule, however, filters which transmit less short-wave radiation are also denser, i.e., they transmit less light generally. Aside from these crude generalizations, there was no principle that the visual scientist could use in evaluating the efficiency of red goggles, and every time the problem arose, different filters had to be compared with each other in a new experiment.

The writer believes that a more universal generalization is possible. The germ of the idea comes from an article by Kohlrausch (**49**) in which he evaluated various filters in terms of what he called the D/T ratio (Dämmerungswert/Tageswert = twilight value/daylight value). The same idea is also used in the article by Hecht and Hsia (**37**). Briefly, the procedure is this: the transmission of the filter in question is multiplied by the relative scotopic luminosity curve. The ratio of the area under the resultant curve to the area under the scotopic luminosity curve yields the scotopic value (S). A similar computation with the photopic luminosity curve yields a photopic value (P). The S/P ratio should be a measure of the relative efficiency of the filter for dark adaptation. The higher the ratio the worse the filter; the lower the ratio, the better the filter. The validity of this proposal has never been thoroughly tested, and it is probably worth investigation. If the principle is valid, it should be possible to work out a table of values showing instantaneous thresholds for the average eye following light adaptation with filters and lights (or lights alone) of various S/P ratios and at various luminance levels. Not only would this eliminate the tedious testing of pairs of filters every time some manufacturing concern proposed a new filter, but it would also provide a rational way of specifying the kind of filter and amount of light (or light alone) which should be used for purposes of dark adaptation.

Research on Methods of Accelerating Dark Adaptation. The flexibility of troops engaged in night maneuvers has always been hampered by the slowness of the dark adaptation process, because military operations do not always occur with enough advance warning to give men time to adapt their eyes. It would obviously be of great tactical value to have some method of accelerating the dark adaptation process.

In recent years a number of articles have appeared from Russian laboratories claiming to show enormous effects of intersensory stimulation on dark adaptation and night vision. Kekcheyev (**46**), for example, reports, "Several months ago we experimented in expediting adaptation by means of light muscular exercise. . . . Experiments made on ten subjects with the help of the adaptometer revealed that it was possible in this way to reduce the period of adaptation from 25–45 minutes to 5–6 minutes." In another article (**45**), the same author claims that "In some instances the period of adaptation dropped from forty-five minutes to eight minutes" with combinations of tastes. Kekcheyev then goes on to say in the same article that auditory, olfactory, gustatory, labyrinthine, thermal, pain, tactile, proprioceptive, and interoceptive stimuli produce changes in the sensitivity of the dark-adapted eye. This is summarized by his statement, "It can be seen, therefore, that excitation of any receptor produces changes in the sensitivity of the dark-adapted eye" (**45**). Other researches show that the "sensitizing" stimulus need not be sensed in order to change the visual threshold. Thus, it appears that ultra-sonic vibrations (**47**), ultraviolet light (**44**), radio waves (**90**), and the geographical latitude of the laboratory where tests are made (**50**), affect rod sensitivity.

These reports are so insistent and the claims so positive that they have attracted some attention in certain quarters. Unfortunately, studies which have been done on this problem in non-Russian laboratories do not confirm the Russian claims. Serrat and Karwosky (**81**), for example, report that auditory stimulation produced no effect on the general or specific color threshold. Thorne (**87**) also reports a few incidental measurements on the effects of auditory stimulation on the just perceptible brightness of flashes of light 0.0001 seconds in duration. His findings appear to show that the effects of auditory stimulation may either increase or decrease the visual threshold. The changes are small, however, and do not exceed 0.3 log units.

The most direct experimental evidence on the validity of the Russian claims comes from experiments by Rose and Schmidt (**76**), Matthews and Luczak (**65**), and Chapanis *et al.* (**9**). Rose and Schmidt studied a great variety of factors—caffeine-metrazol, muscular exercise, strychnine, ultrasonic vibrations, ephedrine, octin, stimulation of taste with saccharin, and Vitamin A in an oil solution and in an emulsion—which are supposed to affect dark adaptation. Their results are well summarized by the statement, "When the experiments were carefully made, and the results analyzed critically by the statistical method, none gave convincing evidence that night vision has been improved or impaired." These findings are at variance with those obtained by Matthews and Luczak, who report that light muscular exercise increases dark adaptation. The magnitude of the effect depends on which of their data are taken as the control data. In any case, the *maximum possible* effect amounted to 0.65 log units at 11 minutes after the onset of dark adaptation and decreased thereafter. Matthews and Luczak conclude that the effects "are not nearly as great as those described by him [Kekcheyev].[1]" The study by Chapanis *et al.* made use of a variety of smells, tastes, sounds, and pressures applied to the back of the hand. These authors conclude that "The results of all experiments are completely negative. None of the stimuli used in this experiment either facilitated or inhibited dark adaptation, contrast sensitivity, or form discrimination at low illuminations."

It is difficult to reconcile these conflicting findings. The situation is especially difficult because the Russian reports are extremely skimpy in presenting data or details of apparatus or experimentation. In their article, Chapanis *et al.* advanced some hypotheses to account for the discrepancies, but the problem cannot be considered settled. It would be highly instructive to obtain a full account of the methods and data of the Russian experiments.[2] Without this information, it appears unlikely that further experimental work on this problem at the present time would contribute anything of value.

The fact that any stimulus—real or imaginary, adequate or non-adequate—seems to produce changes in visual sensitivity in the Russian laboratories makes their results highly suspect. As a purely practical matter it must be recognized that men in military situations are constantly being stimulated by sounds, smells, pressures, internal stimuli, etc. Under these circumstances—assuming for the moment that the Russian findings are valid—one might

[1] Author's insert

[2] It may be of interest to report that the writer attempted to obtain this information in 1943 through military and diplomatic channels. These attempts met with no success. Somewhat more recently (December, 1947), he again attempted to communicate directly with Dr. Kekcheyev via the mails. The letter was returned unopened from Moscow.

expect that the combat soldier or sailor will be at the peak of his seeing ability and that there is little else one can do to help him.

Night Vision

As we have seen above, dark adaptation is usually measured by determining the smallest amount of light an individual can see at various times after he has gone into a dark room. Actually, however, the fighting man is seldom concerned with the mere perception of light, but rather with the recognition and identification of forms. During the war most people assumed that these functions were synonymous and, in fact, this idea is still common even among visual scientists. One important consequence of this belief was that there was a fair amount of muddled thinking about night vision tests. Tests of dark adaptation, contrast sensitivity, form discrimination at low luminances, glare recovery, and many others, were proposed as tests of night vision, and it was generally assumed that these tests all measured the same thing. More recent research has shown that this is not the case, and there has now been accumulated a fair amount of evidence to show that some of the principles of *night vision* may be different from those of *dark adaptation.* It is worth while to review briefly some of these lines of evidence.

Factors in Night Vision. For purposes of this section, night vision will be defined very broadly as the ability to see at rod levels of luminance. A few authorities in the field have generally recognized that, while dark adaptation may set the limit, other factors play an important role in night vision. Thus, Livingstone (**52**) states that "The factors involved in the awareness of the presence of an object and in the acute analysis of the stimulus producing that awareness consist, it seems, of a highly complicated inter-reaction between physiology and psychology." In one of the few experimental studies in this field, Craik and Vernon (**21,** p. 229) conclude:

> The ability of the . . . subjects to perform simple perceptual tasks during dark adaptation, such as reading the position of a dial hand, was very closely related to their absolute thresholds of vision. But the perception of more complicated material, such as silhouette pictures, was determined only partially by the absolute threshold in cone vision; and still less by the absolute threshold in rod vision. The ability of the subject to interpret the vague sensory data obtained in rod vision, and hence to make out the meaning of the picture, was affected by a variety of purely psychological qualities. Intelligence, education, and familiarity with and experience of the particular type of material and situation, were of great assistance. Affectively based attitudes were also important. Too much caution or timidity prevented the necessary amount of guessing at the meaning of the pictures; too much recklessness, and suggestibility to irrelevant influences, produced inaccurate guessing.

About as much as we can say with authority about these other factors is that it is almost certain that perceptual and mental factors contribute to night vision, but that the exact amount of their influence cannot be assessed with our present knowledge.

Data of Night Vision Tests. Data from a large number of scattered wartime reports show that measures of dark adaptation do not correlate with scores on tests involving visual acuity or form perception at rod levels of luminance. Hammond and Lee (**32**), for example, found no correlation between measures of dark adaptation determined with the Hecht-Shlaer Adaptometer and any of their other tests of night vision. Rowland and Rowland (**78**) measured three different aspects of night vision: (a) the light threshold, (b) the minimum intensity of light required for the subject to locate a test object, and (c) the minimum intensity required for form discrimination. The correlations between light thresholds and the other two measures were hardly significant: 0.29 and 0.22. The correlation between the latter two measures, which required the observer to locate or identify a form, or both, was much higher: 0.63. Keil (**43**), Pinson and Chapanis (**74**), Rowland and Mandelbaum (**77**), and Chapanis (**6**) are some of the other authors who have pertinent data on this point.

Since it is not generally realized how many such data exist, it would be a worthwhile project for someone to collect these obscure war-time reports and collate their findings. In many cases, the numbers of subjects were few, but the aggregate findings of these studies constitute legitimate data and their conclusions cannot be ignored.

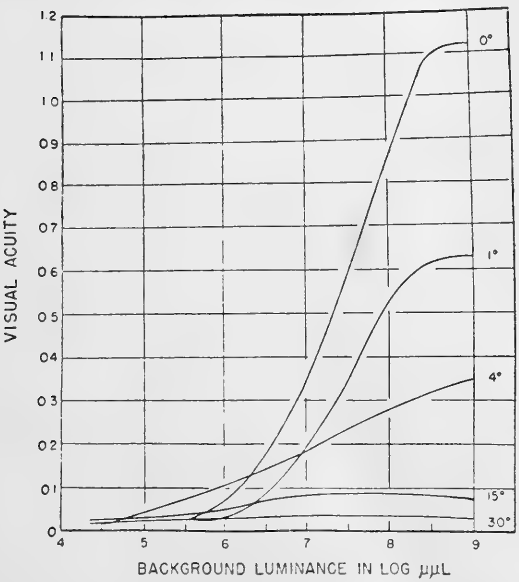

Fig. 15. Visual acuity as a function of luminance at various retinal locations. Notice that at certain low luminance levels, best acuity occurs about 4° from the fovea. (Data from Mandelbaum and Rowland, **64**)

Tests of dark adaptation do not appear to correlate well, if at all, with tests requiring the observer to locate or identify forms at rod levels of luminance. This conclusion is valid for individuals with normal visual capacities. If night-blind or pathological cases are included, the correlations are a little better because of the increased heterogeneity of the data (in a purely statistical sense).

A Given Level of Adaptation Does Not Guarantee Acuity. This point is a difficult one to make because it is almost impossible to find data on it. Briefly, however, it is this: during the course of dark adaptation, when the sensitivity of the eye has increased so that it can perceive light of a certain luminance, it cannot discriminate forms at that luminance level. Adaptation must continue beyond this point before the eye can see things. Miles' data (**66**)—on only one case—illustrate the point. His subject was able to see a surface with a luminance of five log $\mu\mu$L after about 9.5 minutes of dark adaptation, but he could not see a silhouette with a background of the same intensity until six minutes later. Low (**53**) reports confirming data, and concludes that "Maximal form acuity follows light perception by a measurable interval, sometimes as long as 15 minutes. . . . Retinal sensitivity to light of a given intensity does not guarantee maximal visual powers." Low, unfortunately, does not report the intensity of the lights he used, nor is it possible to calculate the luminance levels from the data he gives. This, of course, is a very important requisite for a fuller understanding of the process. Needless to say, further work is needed to elucidate with more precision the relationship between these two aspects of visual performance.

Retinal Variations in Scotopic Visual Acuity. Another very important difference between dark adaptation and night vision concerns the region of maximum sensitivity and visual acuity on the retina. As we have already seen (Figs. 3 and 4), the region of maximum sensitivity to light occurs about 20° from the fovea. This is the basis for the statement, heard many times during the war, that to see things best at night we must not look at them but must look at a point about 20° away from them. Actual experimental data on visual acuity at rod levels, however, show that the region of the retina which is most acute is considerably closer to the fovea. Fig. 15, for example, shows the data obtained in a study by Mandelbaum and Rowland (**64**). These curves show clearly that at certain luminance levels, best visual acuity occurs about 4° from the fovea. Confirming data come from an experiment by Gordon (**30**), who finds the most acute region of the retina at rod levels to be between 5° and 7°.

These data are of practical value, of course, because they show that we see things best at very dim luminances by looking about 5° to one side of them. At rod levels one cannot see things well by looking directly at them. A second, and for theoretical reasons a more important, implication of these data is that they provide additional evidence for the conclusion that dark adaptation and night vision (at least visual acuity at dim luminances) are different functions. The writer can advance no good physiological explanation to account for this state of affairs, but it must be associated in some way with the rod population and neural interconnections we spoke about earlier.

Conclusions. A general conclusion which can be drawn with some assurance from this evidence is that night vision is a complex visual process. Dark adaptation is one of the factors which contributes to night vision but there are other factors which may be as important, or more important, in determining night visual capacity. What these other factors are, and how much they contribute to night vision, is not completely understood. The fact that dark adaptation and night vision appear to be largely independent is of practical importance not only in designing night vision tests, but also in instructing men engaged in night duties.

Visual Acuity

The Measurement of Visual Acuity

Visual acuity, in its broadest definition, means the ability to distinguish fine detail. In actual practice there are many different ways in which acuity is measured and expressed. We shall discuss some of the more common techniques briefly here.

Snellen Ratings of Acuity. Perhaps the most common and familiar method of testing visual acuity is by means of a doctor's eye chart. The typical chart of this sort contains a series of letters of various sizes, and the examinee reads as many as he can. The smallest detail an observer can see is a measure of his acuity (see Fig. 16). When acuity is measured in this way it is usually expressed as a ratio of the distance at which a given line of letters can be seen by the examinee to the distance at which a person with normal vision can see the same letters. Thus, a rating of 20/20 means that the examinee could see at 20 feet letters which the normal person can read at 20 feet. A rating of 20/40 means the examinee could just read at 20 feet what the normal person can see at 40 feet. A rating of 20/10 means that the examinee could see at 20 feet what the average person can just barely see at 10 feet.

Actually, these ratings are ambiguous because they depend greatly on the definition of what a "normal" person can see. A rating of 20/20 is usually taken as the normal rating, but, as we shall see, this varies greatly depending on the amount of illuminance, the contrast between the object and its background, etc.

Two other kinds of visual acuity test objects are also shown in Fig. 16. These are the Landolt ring and parallel bars. Both of these test objects can be used with illiterates, of course, and so have an advantage over the Snellen letters for such subjects. A much more important advantage, however, is that they largely eliminate a form perception factor which vitiates results obtained with letter tests (**103**). Even though a series of letters may have the same overall size, the same thickness of lines, and so on, some of them will be more legible than others simply because of their form; e.g., the C, O, and G are easily confused whereas the T and L are fairly distinctive. The Landolt ring and parallel bars eliminate this perceptual factor by keeping the form of the test object constant. The test factor, the position of the break in the Landolt ring (or the orientation of the parallel bars), is changed from trial to trial. A series of rings or bars, varying in size, is used to measure acuity. Many other kinds of test objects have been used to measure acuity, and some of them are described in a report of the Personnel Research Section of the Adjutant General's Office (**103**).

Visual Angle. An index or measure of visual acuity which the laboratory scientist prefers to the $20/x$ notation is the size of the smallest detail an observer can see in terms of the visual angle subtended by this object at the eye (see Fig. 16). This kind of measure is convenient because it makes no assumptions about the visual acuity of a hypothetical normal person, and it gives us a single number to work with. Snellen ratings of acuity are usually translated into measures of acuity expressed in visual angles by an arbitrary conversion factor, viz., a Snellen rating of 20/20 corresponds to a visual angle of one minute.

Acuity as measured by the test forms in Fig. 16 does not represent the maximum acuity of the eye. Recent measurements show, for example, that under ideal conditions the eye can detect the presence of an object which subtends about a half second of visual angle (**38**). This is roughly equivalent to seeing a wire, $\frac{1}{16}$th inch in diameter, a half mile away! This, however, is more a measure of visibility—about which we shall say more later—than of visual acuity.

Physiological Factors Affecting Acuity

Near and Far Acuity. A very important factor which determines visual acuity is the distance at which the test is made. At first glance, this might not appear reasonable if the visual acuity is expressed in angular measurements. Actually, experimental evidence shows that visual acuity at 13 inches may be quite different from acuity at 20 feet (**29, 85**). Acuity for distances greater than 20 feet appears to remain fairly constant. The reason for this is that the lens in the front of the eye has to change shape in order to focus near and far objects upon the retina (accommodation), and people differ markedly in their ability to do this. Accommodation remains fairly constant for objects at distances greater than 20 feet.

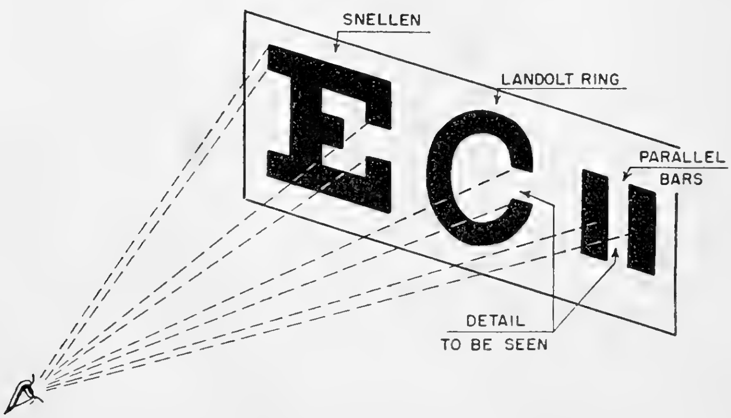

Fig. 16. Three types of test object used in the measurement of visual acuity. In the Snellen letter there are actually many critical details to be seen. (After Crouch, **23**)

This fact is of some practical importance. In some industries, and in many military situations, men are selected for particular jobs on the basis of results obtained in physical examinations. Because visual acuity is related to accident-proneness and efficiency in certain types of jobs, this is a critical item in the physical examination. This means, then, that if a job calls for good distance acuity, e.g., truck driving, the men should be tested at 20 feet. But if the job calls for good near acuity, e.g., toolmaking or operating a radar, then the men should be tested at a distance comparable to that used on the job. Because all other factors which affect acuity seem to affect near and far acuity alike, we shall not have to talk more about distances in the following discussion.

Visual Acuity and Retinal Location. We all use our eyes so much that we commonly overlook certain peculiarities about our seeing process and would never notice them if they were not pointed out to us. If, for example, you stare steadily at a letter on this page, it is impossible for you to read letters two inches away; or if you look at the road straight ahead, while driving, you will not be able to read most signs along the side of the road. The variation of visual acuity in different parts of the eye is plotted in Fig. 17. In this illustration, zero degrees represents the straight-ahead direction—sometimes called the line of sight. Light rays from this direction strike the fovea—the center part of the eye. Notice that visual acuity drops off rapidly for objects in the periphery of the eye. As a matter of fact, when you look straight ahead, your visual acuity five degrees to the right or left of the central line of sight is just half as good as it is in the fovea. At 40, 45, and 50 degrees from the central line of sight, visual acuity is only about $\frac{1}{20}$ of what it is directly straight ahead. It is interesting to note, incidentally, how closely this curve parallels the distribution of cones in different parts of the retina (see Fig. 2).

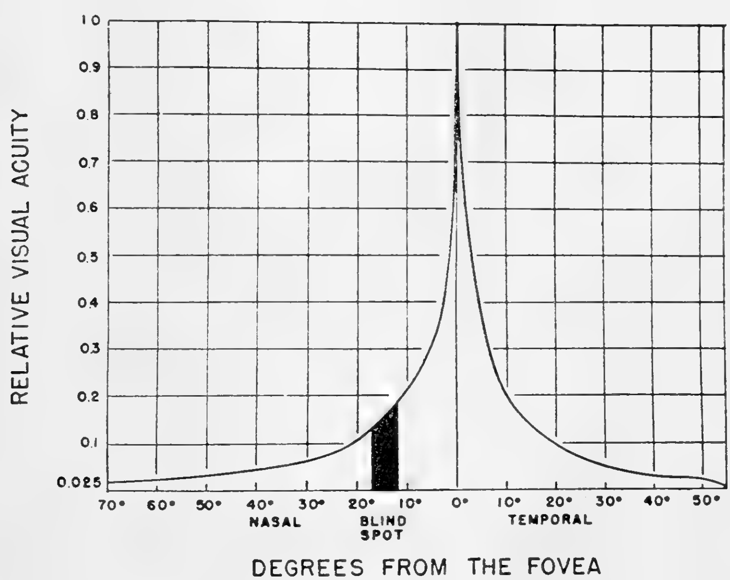

Fig. 17. Visual acuity at different retinal positions. (After Wertheim, **97**)

The Blind Spot. One thing more before we leave Fig. 17. Notice that there is a blacked-off area in the figure. This is labelled the "blind spot." This area was also marked off in other figures we have used so far (see Figs. 3 and 4). The blind spot is the place where the nerves and blood vessels come into the eyeball. It is located about 15 degrees from the fovea, on the side of the retina nearer the nose, and is about seven degrees wide and about five degrees high.

Since there are no rods or cones in the blind spot, it is literally blind. You cannot see anything in this area; and yet if you try to find a hole somewhere in your visual field, you probably cannot find one. It is there, nonetheless, and any laboratory psychologist could demonstrate it very simply.

Physical Factors Affecting Visual Acuity

In contrast to the previous section, our concern here is how visual acuity is influenced by physical factors in the environment. There are so many research studies on these factors that it would be impossible to review them all here. It has been necessary, therefore, to select representative data to illustrate each point.

Luminance and Visual Acuity. We have already had occasion to note the importance of the first factor, viz., luminance (see Fig. 15). The curve in Fig. 18 is an average curve drawn through the data of six thorough investigations on visual acuity as a function of background luminance. For values above the cone threshold, visual acuity was measured for foveal or central vision. This curve illustrates several important points. First, notice how rapidly visual acuity increases as the luminance increases in the middle range of values—from 0.01 to 100 mL (log = −2 to log = 2). Beyond this value visual acuity does not increase so much, although the curve continues to rise indefinitely, even if very slowly, as the luminance increases.

Although the curve in Fig. 18 is the kind frequently used in discussions on vision, it is important to note that there are alternative ways of plotting the same data and that these alternative methods give slightly different results. The scale at the left in Fig. 18 is a scale of visual acuity, defined as the reciprocal of visual angle in minutes. If one compares this scale with the visual angle scale at the right in the figure, it becomes immediately apparent that the visual acuity measure, by its very nature, yields a distorted scale. A change in visual acuity from 0.5 to 1.0 means a change of one minute in the size of the smallest discriminable object, i.e., from 2.0 to 1.0 minutes. A similar change

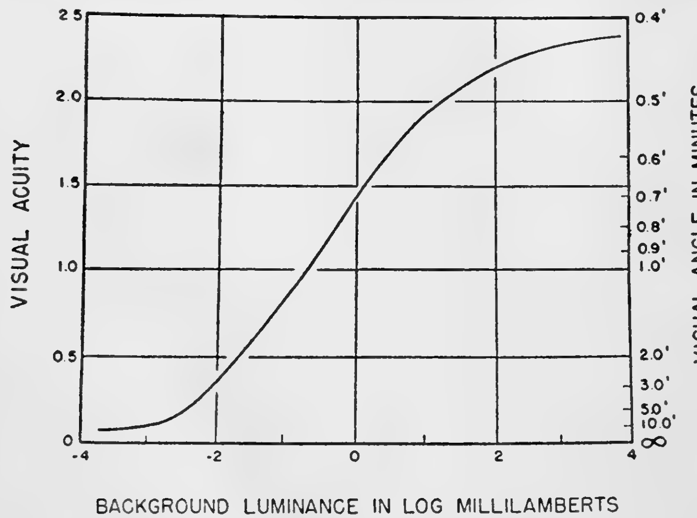

Fig. 18. Visual acuity as a function of background luminance. (From data compiled by Moon and Spencer, **69**)

in visual acuity from 2.0 to 2.5, on the other hand, represents a change in object size of only 0.1 minute. In terms of the actual size of the object to be seen, then, the visual acuity scale greatly expands differences between the smaller objects and contracts differences between the larger ones.

The curve in Fig. 19 shows the same data as those in Fig. 18. In Fig. 19, however, we have plotted the actual size of the smallest detail which can be seen. Although the two curves plot the same data, what they show is quite different. The curve in Fig. 18 appears to show large increases in visual acuity when the background luminance is increased from 10 to 1000 (log = 1 to log = 3) mL. Fig. 19, on the other hand, shows that in terms of the smallest detail which can be seen, there is not much to be gained by increasing the luminance within this range. It is important to note, incidentally, that the ordinate in Fig. 19 is logarithmic. Thus, even in this figure the high values are compressed and the low ones expanded. For most applied visual problems, it is likely that the kind of plot illustrated in Fig. 19 is more meaningful than the kind illustrated in Fig. 18.

Visual Acuity at Night. These data are also important for telling us about seeing at low luminance levels. The decrease in acuity at night means that identification of objects must depend on the perception of generalized contours and outlines and not of small distinguishing features. Wires, picket fences, and telephone poles may be invisible a few hundred feet away at night. Aircraft and ships are least visible when viewed from dead astern, because their areas are smallest in that direction. For this reason, night interception tactics during the war required that enemy aircraft be followed from rear-above or rear-below, rather than rear-level. Similarly, small terrain features, a small building, smokestack, or a bridge, may not be visible from the air at night, and recognition must depend on rather large ground features—surf and sand, large clumps of trees, rivers, lakes, and large concrete installations.

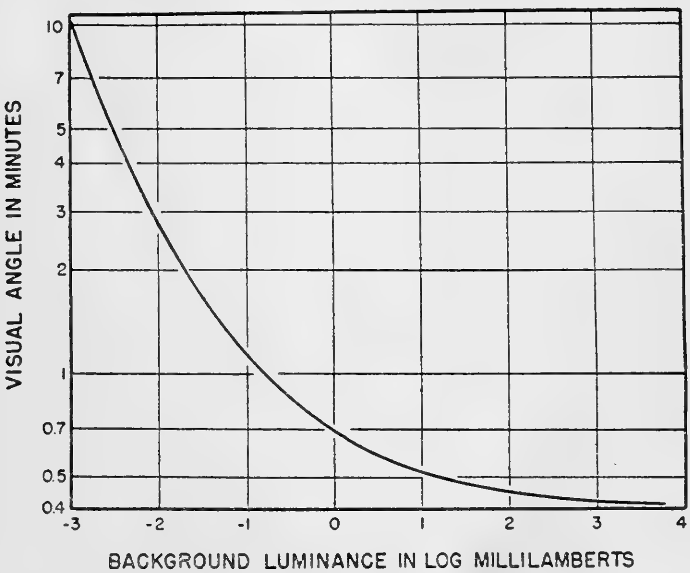

Fig. 19. This is the same curve as in Fig. 18, except that the ordinate is now a logarithmic plot of the smallest detail visible. Notice that this curve shows lesser gains as luminance is increased at the higher values.

Time and Visual Acuity. The second

factor affecting visual acuity is time: an object is more visible the longer you look at it. The data in Fig. 20 are taken from an experiment by Ferree and Rand (**24**). In this experiment, targets of various sizes, illuminated to various luminance levels, were exposed for very short periods of time. The observer's task was to identify the target during the short exposure. The data in Fig. 20 are plotted in terms of speed, i.e., the reciprocal of time in seconds. It is evident that when the subjects had

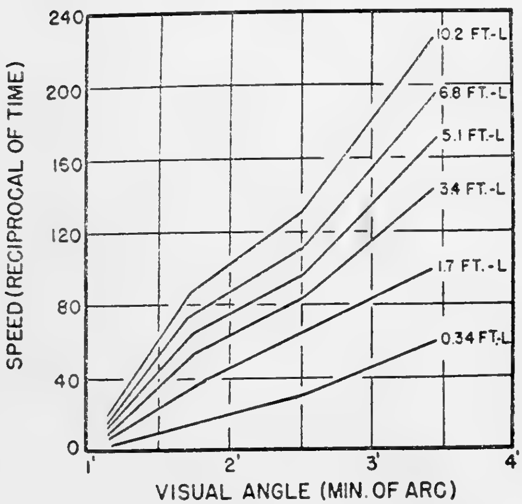

Fig. 20. At any luminance level, less time is required to see bigger objects. When size is held constant, less time is required to see at higher luminance levels. (After Ferree and Rand, **24**)

more time to look at the target, they could see much smaller objects. Also evident is the influence of the first factor—luminance. At any exposure time, the subjects were able to see smaller targets as the background luminance increased.

The range of background luminances covered in this experiment is very small, and there do not appear to be any more complete studies of this function. There is good reason to believe, however, that time would become an especially critical factor at luminance levels below the cone level. It would be most instructive to have data for this function over a much greater range of background luminances.

It should be noted in passing that the use of a "speed" scale in Fig. 20 is open to the same objection that we had about the visual acuity scale in Fig. 18. Speed, like visual acuity, is a reciprocal measure which distorts the original experimental measures. Differences in the actual times represented by speed values of 160 and 200 are so small, 0.006 as compared with 0.005 seconds, as to be practically insignificant. Reconverting into time measures would not change the essential correctness of our conclusions about Fig. 20, but it would change our estimates about the practical importance of some of the trends shown here.

Visual Acuity and Luminance Contrast. The term luminance contrast may be defined by the following equation:

$$C_L = \frac{|L_0 - L_B|}{L_B}.$$

L_O is the luminance of the object, L_B is the luminance of the background on which the object is located, and C_L is the luminance contrast in percent. This term is most frequently referred to as "brightness contrast," but the latter, in addition to being technically incorrect, is confusing because it has a special meaning for psychologists. Psychologists think of brightness contrast as the phenomenon of simultaneous brightness contrast (see Appendix to this chapter).

Basic data showing the relationship between visual acuity and luminance contrast are contained in Fig. 21. The two parts of the figure were obtained from two different studies, and, in general, they agree fairly well for the lower contrast values but do not agree for the high contrasts. These discrepancies are due to differences in the experimental conditions used in the two studies. The data on the left were obtained with long exposure times: the subjects were allowed to look at the targets for three seconds or more. The data on the right were obtained with brief exposures: 0.17 seconds. Since we have already seen that visual acuity is a function of time, it is not surprising to find that the two sets of data do not agree.

Interrelationships Between Factors. Each of these factors in visual acuity is related to every other one. A reduction in any one factor—background luminance, size, contrast, or time—may be compensated for by an increase in one or more of the others. For example, an object which is so small that it is just below threshold may be made visible by increasing the background luminance, time of exposure, or luminance contrast. This is true within certain limits, of course, because it is possible to find an object so small that it cannot be seen by the unaided eye under any circumstances.

The interrelationships between these factors have been investigated by Cobb and Moss (**15**) and have been summarized in graphical form by Luckiesh and Moss (**58**) in Fig. 22. The area above the three-dimensional curve shows those objects which can be seen clearly; the area below the curve represents those objects which cannot be seen. This plot was derived from over 100,000 separate measurements, and it forms the basis for many of our predictions about visual acuity under many practical conditions. This set of data does not cover the full gamut of luminance values, and it has been worked out for only two exposure times. However, it is the best set of data we have on this problem, and the more complete investigation of the interrelationships of these factors must wait upon further research.

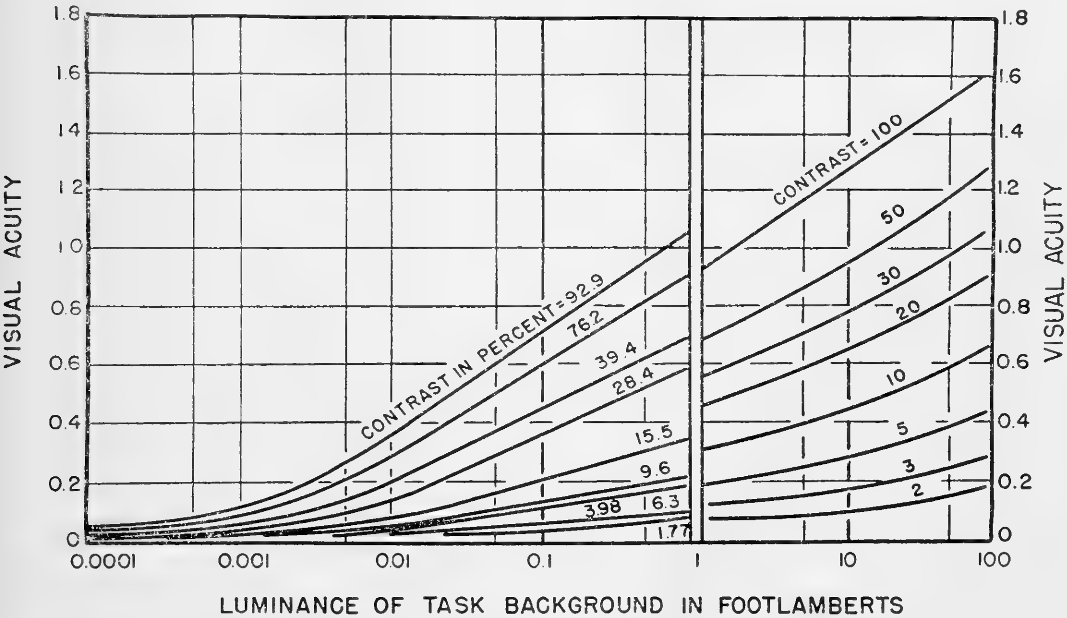

Fig. 21. Visual acuity as a function of background luminance and the luminance contrast between the object and its background. The data on the left are from Connor and Ganoung (**20**); those on the right from Cobb and Moss (**15**).

Luminance of Surround. According to many illuminating engineers, another important factor influencing visual acuity is the luminance of the large area surrounding the visual task. The visual task usually refers to a small area (5° or less) in the center of the field of view and the surround to the much larger area around the visual task. Actually, neither area has been defined precisely in experimental work. Nonetheless, it is generally agreed that visual acuity in the center of the visual field may be changed by varying the luminance of the surrounds. The results of a study by Lythgoe (**62**) on this function are shown in Figs. 23 and 24. These figures suggest that acuity is best when the surround has about the same luminance as the background immediately around the visual task. As a practical illustration of this principle, Hanes and Williams (**34**) have shown recently that detectability of targets on radar scopes actually improves if the illumination in the radar room is kept a little higher than is customary.

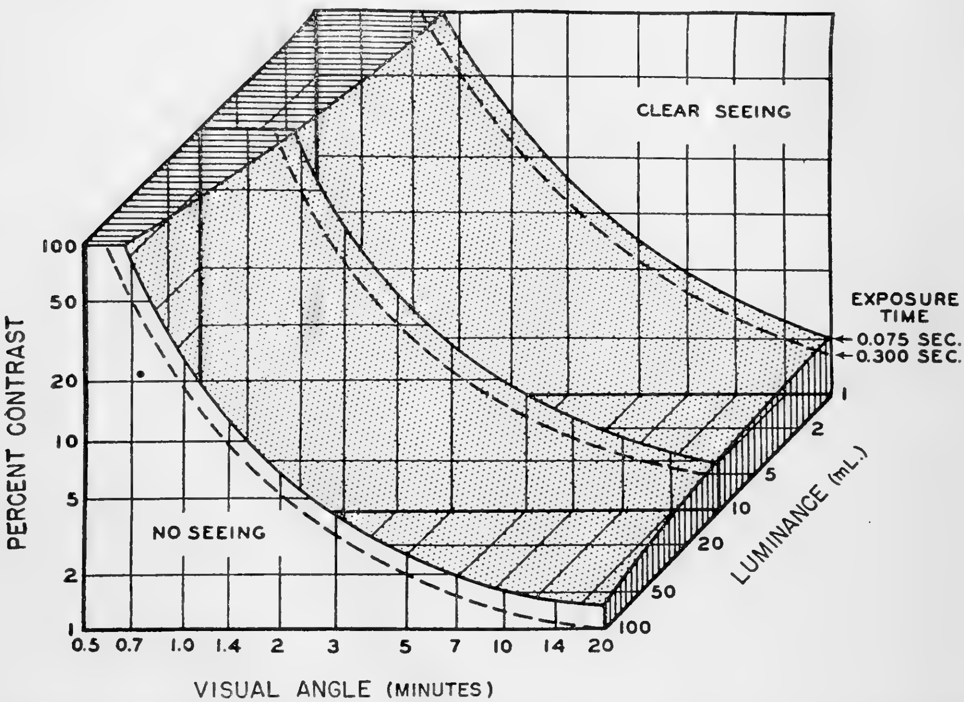

Fig. 22. The interrelationships between visual acuity, background luminance, luminance contrast, and duration of exposure. (Data from Cobb and Moss (**15**) as replotted by Luckiesh and Moss, **58**)

Although these are informative studies, they provide a good illustration of one of the difficulties we mentioned earlier, viz., incompleteness. The study by Lythgoe (**62**), for example, investigated the complete visual acuity function for only three surround luminances (Fig. 23), and his is probably the most thorough study of these functions. In addition, the data in Fig. 23 are for only two subjects; those in Fig. 24 for three subjects. Further, there are very few observational points in either figure. This is an especially critical matter when one comes to draw a curve through the observations. The curve drawn in Fig. 24 is mine, and it shows best acuity to occur when the surround has a luminance of about five equivalent foot-candles. When Lythgoe drew a curve through the same data, he made the curve reach its maximum much further to the left, i.e., at a much lower surround luminance. What one does with data of this sort is somewhat arbitrary, and a maximum could be precisely located only if more data were available. Lythgoe, finally, does not state what the exact size

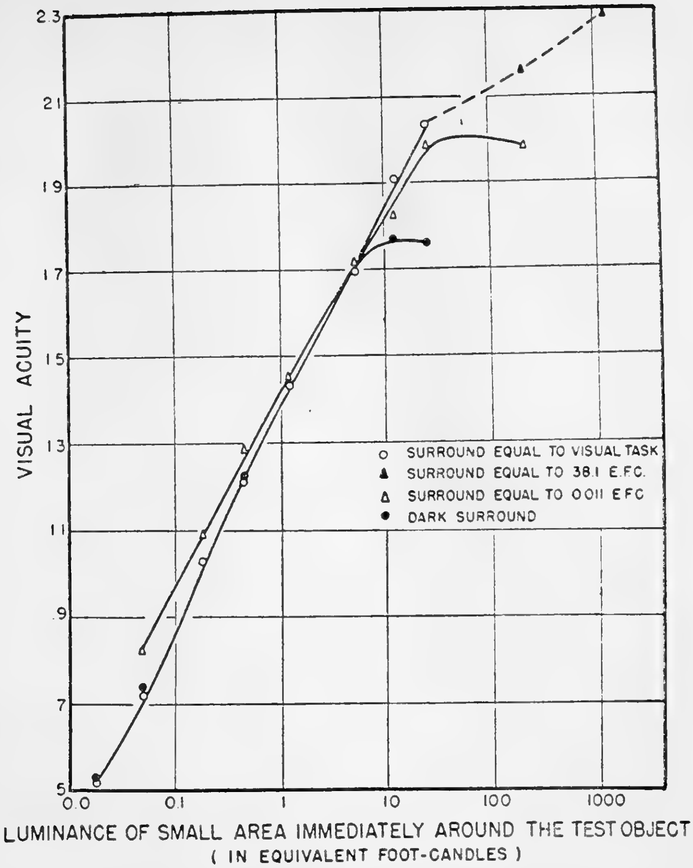

Fig. 23. Visual acuity as a function of the luminance of the small area immediately around the test object with three different surround luminances. The luminance of the surround could not be made to match that of the task at the two highest luminance levels. These are average data for two subjects. (Data from Lythgoe, **62**)

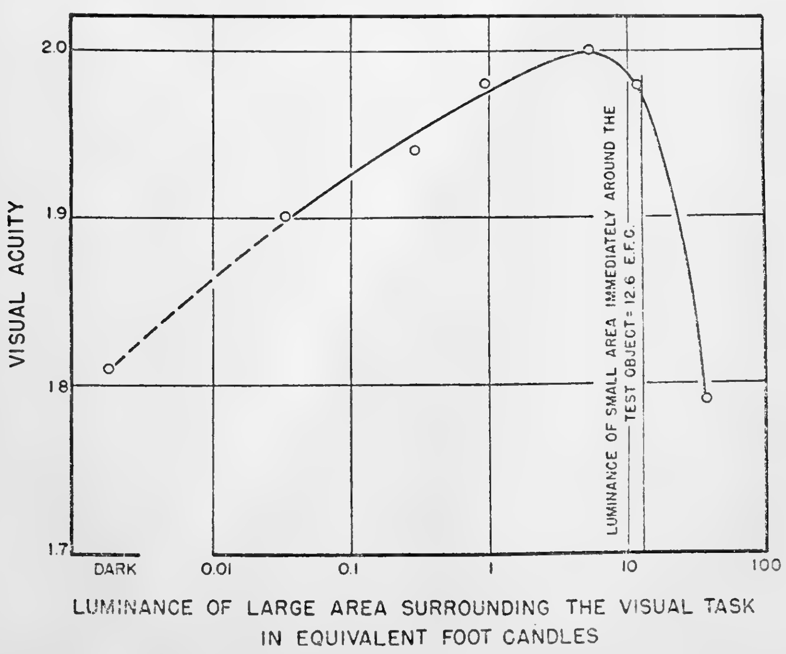

Fig. 24. Visual acuity as a function of surround luminance with the luminance of the small area around the visual task held constant. These are average data for three subjects. (After Lythgoe, **62**)

of the central visual task was, nor is it possible to calculate this value from the information he gives. It appears probable, however, that the central task area was about one or two degrees in size.

These considerations do not seem to have deterred illuminating engineers from writing specifications for surround luminances. Thus, Luckiesh (**57**) says:

> 1. Brightness-ratios smaller than 1 to 5 are desirable.
> 2. Brightness-ratios greater than 1 to 10 should be avoided if reasonably possible.
> 3. Brightness-ratios of 1 to 100 should not be tolerated.

The evidence he adduces for this specification comes from a study in which he investigated the minimal perceptible luminance difference as a function of three surround luminances: (a) five times that of the task, (b) equal to the task, and (c) $\frac{1}{5}$ that of the task. Needless to say, it is almost impossible to determine maxima or minima in any set of experimental results that contain only three observed points. In another experiment, he studied rate of eye blinking as a function of five surround luminances. Since it is questionable whether the rate of eye blinking has much to do with seeing (**1, 88**), these data may be discounted.

Another engineering authority, Crouch (**23**), states categorically, "Thus he [Lythgoe] proves that the best acuity is to be expected only when the surrounding brightness is equal to that of the task." Lythgoe, however, concluded that best visual acuity occurs when the surround is $\frac{1}{100}$ to $\frac{1}{10}$ that of the visual task. The data in both Figs. 23 and 24 probably support Lythgoe better than Crouch. Fig. 23 shows, as a matter of fact, that throughout most of the range of task luminances (a) the dark surrounds and equality surrounds have about the same effect on acuity and (b) acuity is best for the dimly illuminated surround, i.e., 0.011 equivalent foot-candles. It is worth pointing out, incidentally, that Fig. 23 shows the complete data from Lythgoe's report. In his monograph, Lythgoe drew up (in his Fig. 10) the data for only one of the subjects—the one which gave the neater functions—and reversed two of the points at that. Further, Lythgoe did not plot complete functions but only showed the data for the upper range of luminances (from 1.21 equivalent foot-candles up). In spite of these criticisms, authoritative publications by illuminating engineers (**23, 99**) have cited the incomplete and incorrect Fig. 10 from Lythgoe's original paper.

Other studies, occasionally referred to in secondary sources, also turn out to be of little value to the critical scientist. Two papers by Cobb and Moss (**13, 14**), for example, report data on only two surround conditions: dark surrounds and surrounds equal in luminance to the central task. Although visual performance was better with the equality surround, this obviously does not help us decide whether some other surround condition, e.g., a surround equal to $\frac{1}{10}$ that of the task, as suggested by Lythgoe, might not turn out better. A study by Cobb and Geissler (**12**) studied only dark surrounds and surrounds equal to 42 candles per square meter. Another study by Cobb (**10**) reports data for dark surrounds and surrounds equal to 2.9 candles per square meter. Both are obviously very incomplete and the conclusions made by the authors are cautiously worded. Thus, Cobb (**10**) states, "Surroundings of a brightness about equal to or less than that of the test object show no consistently better or worse results than dark surroundings with the identical test object." Johnson has reported an extensive investigation (**41**), not on visual acuity but on reaction time, with task luminance held constant and with six ratios of task/surround luminances. If one selects data for identical subjects under all conditions, his results show that visual performance is markedly poorer when the surround luminance is *greater* than the central area. When the surround luminance is less than that of the task, visual performance is not consistently affected one way or the other. Cobb's data (**11**) on luminance sensitivity—the smallest luminance difference which can be seen—are in essential agreement with those of Johnson, except that Cobb appears to find a somewhat more consistent, though still very small, decrease in visual performance for surrounds which have a luminance lower than that of the task. Although raw data are not given, the present writer questions whether the differences are statistically significant.

Fisher's study (**27**) is interesting because it attempted to provide quantitative data on the influence of the size of the surrounds. His central task area had a constant diameter of 2°. Five different surround sizes were used: 2.5°, 5°, 7.5°, 12.5°, and 20° in radial width measured beyond the edge of the central task area. His results show that when the surrounds are brighter than the task, acuity decreases with increasing size of the surrounds. When the surrounds are dimmer than the central task, acuity increases with increasing size of surrounds. When the surrounds are equal to the central task in luminance, increasing the size of the surrounds has no consistent effect on acuity. It should be noted, incidentally, that if his data are analyzed in terms of the task/surround luminance ratio, in only a few instances was acuity best when the surround and task luminances were equal.

This problem is discussed here in some detail because it is a good illustration of the need for more complete studies of many visual functions. Once specifications have been prepared for illuminances to be used in various seeing situations, they have an air of finality about them. The reader gets the impression that the specifications stem from sound experimental evidence. It is rather startling at times, therefore, to discover how meager the evidence really is.

Chromatic Contrast. A factor which is largely ignored in studies of visual acuity is chromatic contrast. Walls (**93,** p. 496) puts the situation in an amusing fashion:

> A gentle indictment can be laid against all the investigators who have wrestled with the problems of resolution and come away from the strife proudly clasping theories to their bosoms. One and all, they seem to be color-blind.

The gap in our knowledge on this point is serious. One gets the impression from reading most authors that luminance and luminance contrast are the really important factors in determining visual acuity. But it is easy to demonstrate that visual acuity is possible when a red object is placed on a green background of the same luminance.

A start was made on this problem during the war in a research contract held by the Eastman Kodak Company (**101**). It was admittedly only a beginning and this work needs to be pursued further. A tentative conclusion which came out of the Eastman work was that luminance contrast and chromatic contrast both contribute to visual acuity according to this equation:

$$C_E = \sqrt{C_L^2 + C_C^2},$$

where C_E is the effective total contrast, C_L is the luminance contrast, and C_C is the chromatic contrast. Whether this generalization will be valid when more data are accumulated is difficult to say. One thing is clear, however; if we do not take chromatic contrast into account, we are probably ignoring an important determinant of visual acuity.

Visual Acuity as a Function of Color of Illumination. A special case of the problem of chromatic contrast concerns visual acuity as a function of the color of illumination. Usually, the problem is one involving black or gray test objects seen against a white background illuminated with different colored lights. This means, in general, that the background takes on the color of the light, so that chromatic contrast varies. In addition, the reflectance of the black test object is usually much less than that of the background, so that there is some luminance contrast present also.

One set of results obtained on this function is illustrated in Fig. 25. These curves are taken from a study by Ferree and Rand (**25, 26**) and represent average data for three subjects. In one-half of the experiment, the visual acuity test was illuminated

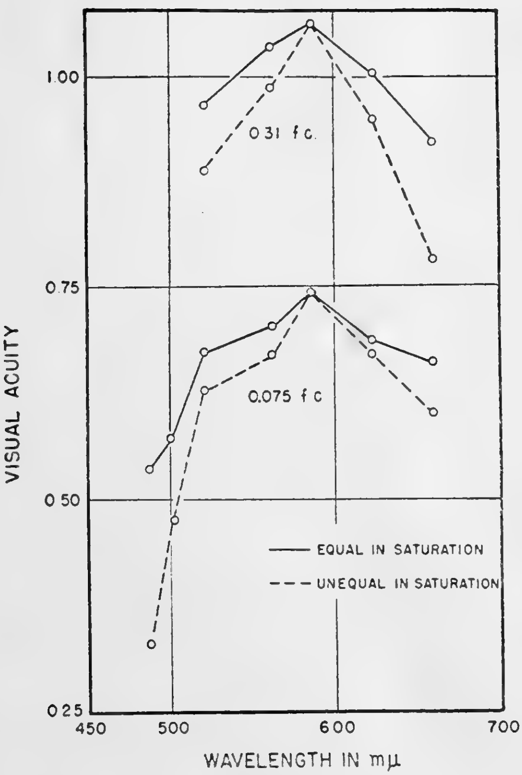

Fig. 25. Visual acuity as a function of the color of illuminance. (Data from Ferree and Rand, **25, 26**)

with spectrally pure light of the wavelengths shown. Spectrum lights, however, differ markedly in both dominant wavelength and colorimetric purity (see Appendix) and in their psychological equivalents, hue and saturation. To isolate the effect of hue alone in the second part of the experiment, these investigators equated the saturations of the illuminants by mixing enough white light with the spectrum lights until they appeared equally saturated. The results from both parts of the experiment agree in showing that visual acuity is best when the illuminant is yellow in color. Visual acuity drops off markedly when the colors are selected from the ends of the spectrum. Almost identical results were obtained in a study by Luckiesh (**56**), also done apparently on three subjects.

It is difficult to get very much intensity from most optical systems used to produce spectral lights. For this reason, the illuminances used in the study by Ferree and Rand were very low: 0.075 and 0.31 foot-candles. Since the reflectance of the white background was about 85 percent, the background luminances were about 0.064 and 0.26 ft.-L, respectively. In the study by Luckiesh, the illuminance was 4.2 foot-candles but the background luminance cannot be computed from the data given. This is less light than is ordinarily used for reading and other visual tasks. It would be informative to have similar experiments done with much higher light values.

Two other findings of the Ferree-Rand study are worth noting. (a) Measurements of speed of vision (time needed to recognize a test object) and the power to sustain clear seeing (the length of time an observer could focus the pattern "li" clearly) gave results essentially identical to those obtained with the acuity measurements. For all three visual functions, yellow light gave the best results. (b) Light from a Macbeth daylight lamp (roughly equivalent to illuminant C) gave better results than any of the spectrum lights even when they were all equated in illuminance. Light from a Mazda lamp—which is considerably yellower than a daylight lamp—gave results almost the same as those obtained with spectral yellow.

Visual Acuity as a Function of Color of Illumination at Rod Levels. Two experiments of a slightly different sort are those reported by Shlaer *et al.* (**82**) and Luckiesh and Taylor (**59**). Shlaer and his collaborators studied visual acuity over a wide range of luminances, but with only two colors of illumination, red (670 mμ), and blue (490 mμ). Their results, for one of the two subjects they used, are shown in Fig. 26. Notice in particular the data at the low luminance values, which show visual acuity to be much better with blue light than with red light. Exactly opposite results have been reported by Luckiesh and Taylor (see Fig. 27). Since the data by Luckiesh and Taylor do not cover as great a range as those of Shlaer *et al.*, the scale in Fig. 27 is slightly expanded. Further, in reworking both sets of data to make them comparable, the writer had to make certain assumptions which may or may not be valid. In any case, the data within Figs. 22 and 23 are strictly comparable, even though the data of Fig. 22 may not be strictly comparable to those in Fig. 23.

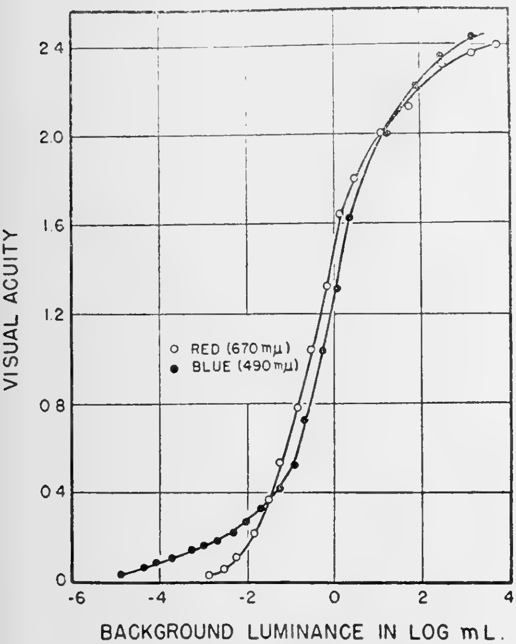

Fig. 26. Visual acuity with red and blue backgrounds of various luminances. (Data from Shlaer *et al.*, **82**)

Notice particularly that Luckiesh and Taylor find visual acuity under red light at low luminances to be far superior to visual acuity under blue light. The resolution of this apparently complete contradiction of results probably lies in the way the two groups of experimenters carried out their photometric measurements at the low luminance levels. Shlaer *et al.* do not say explicitly how they made these measurements, but the implication is that they used a technique similar to that used by Blanchard (**3**) about which we commented earlier. They equated luminances at high levels and computed the luminances for the rod levels by means of the densities of the neutral filters they used. In fact, in one place in their paper they say, "The intensity scale is, by definition, that of the cones." This is another way of saying that at −2.5 log mL, for example, the blue light probably was much brighter than the red light. Luckiesh and Taylor, on the other hand, probably used a low-brightness photometer in measuring luminance values. This is another way of saying that the luminances at each level were adjusted until they appeared equally bright.

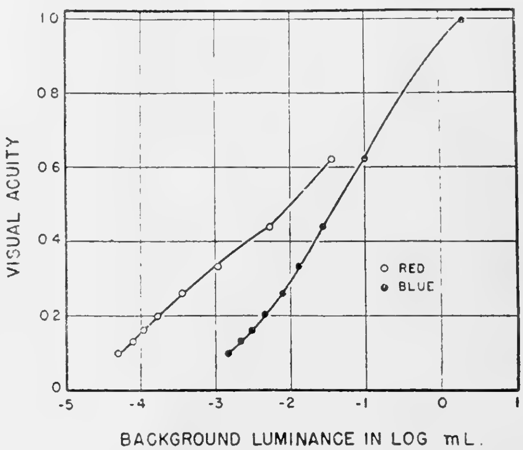

Fig. 27. Visual acuity with red and blue backgrounds of various luminances. (Data from Luckiesh and Taylor, **59**)

We cannot be sure, of course, that this explanation accounts for the entire discrepancy between the results of these two experiments, but it undoubtedly accounts for most of it. If this analysis is correct, the results obtained by Luckiesh and Taylor are probably the more meaningful for most applied seeing problems. When lights are adjusted to appear of equal luminance to the eye, visual acuity at very low luminances is better under red light than under blue light. Further work is needed to discover whether this interpretation by the present writer is correct. In any case, this is another good example of the kind of inconsistency which crops up in visual data because of the lack of a consistent system of low-luminance photometry. Different writers do not always mean the same thing when they report their measurements for visual functions in the region of rod sensitivity.

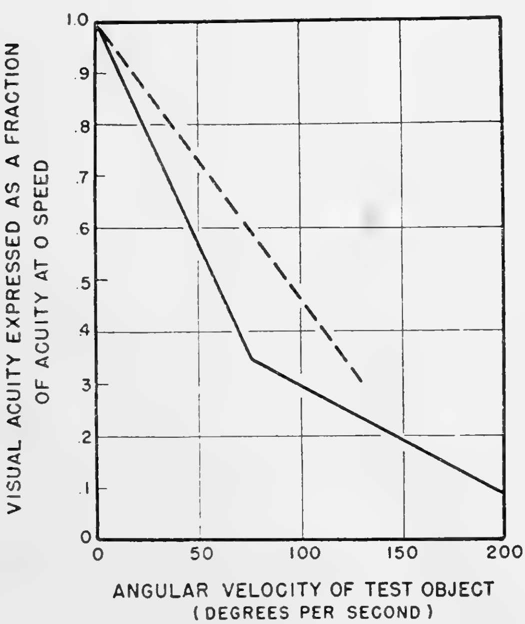

Fig. 28. Visual acuity for objects in motion. (Data from Ludvigh, **60, 61**)

Visual Acuity for Moving Objects. Most visual acuity studies have been done with a stationary observer and stationary test object. Yet many practical seeing situations involve motion, either of the observer or of the object being observed. It is important in this connection to make a clear distinction between the perception of motion and visual acuity for moving objects. Perception of motion means simply that an observer has seen something move; visual acuity for moving objects implies that the observer can recognize or identify critical details of the object in motion.

A few studies which have been done in this area show conclusively that visual acuity is greatly impaired if an observer is looking at a moving object. Two sets of data obtained by Ludvigh (**60, 61**) are shown in Fig. 28. They are for monocular photopic vision and for a constant angular velocity of the test object in a horizontal plane around the observer's head. The earlier data (interrupted line) were obtained with Snellen letters; the later data (solid lines) with Landolt rings. Low (**54**) has investigated visual acuity for moving objects in the periphery of the eye. He used Landolt rings, with apertures varying from 3′27″ to 2°17′27″, moving at a constant velocity of approximately 15° per second. His results show that visual acuity for moving objects seen peripherally is only about 60 percent as good as visual acuity for stationary objects seen in about the same location. His results, however, are essentially complete only for one position: a point 30° from the fovea. At 45° from the fovea, 32 out of 200 observations were incomplete because subjects could not correctly identify even the largest test object; at 60°, 144 out of 200 observations were incomplete for the same reason.

This brief summary indicates clearly that here is an area which has scarcely been touched by visual research. Although it is quite likely that they will not yield results of great theoretical importance, additional data on these functions might be of some practical value in the solution of problems dealing with the visibility of rapidly moving targets.

Visibility

For purposes of discussion here and in the chapters to follow, visibility means recognizing the presence of a light or object without having to recognize its shape or form. In this sense it is, perhaps, synonymous with detectability. This is not the sense in which Luckiesh (**57**) uses the term visibility, but it is more in accord with recent usage (**2**). In general, there should be some kind of relation between visual acuity and visibility, and we might expect visibility to be low order visual acuity. This is essentially correct. In addition, we find that the same factors that affect visual acuity—luminance contrast, time, size, and luminance—also affect visibility. It should also be clear that much of the data presented in the section on rod and cone vision are visibility data. This section, however, will consider in slightly more detail some basic visibility data which we shall make use of in Chapters 3 and 6.

Visibility of Lights

Colored Signal Lights. Many light-signalling devices require us to recognize the colors of lights. In a general way, we know that if we are to do this, the amount of light reaching the eye must be greater than is required for merely telling the presence of the light. Further, the more complicated the color system is, the higher must be the amount of light for each color. If one must recognize three different colors in a light signalling system, more light is required than one would need to recognize two. Although these generalizations are undoubtedly valid, exact quantitative data for putting them to practical use are not available.

There are available, however, some very informative experiments by Hill (**40**) on this problem. He used a point source of light, against a background equal to a starlit sky, with exposures of 1.5 seconds, and with two intensities of light—one which produced one mile-candle of illuminance at the eye, and another which produced two mile-candles. The observers knew where the light was and looked in that direction. Over 30,000 observations were made with 73 different colored lights which the observers had to identify as red, yellow, orange, white, green, or blue.

The results of the experiment show that blue and green (and yellow and orange) are indistinguishable when they are far away. Red and green (or blue) were the easiest to recognize, then came white and last yellow (or orange). A portion of the quantitative data obtained in this study are shown in

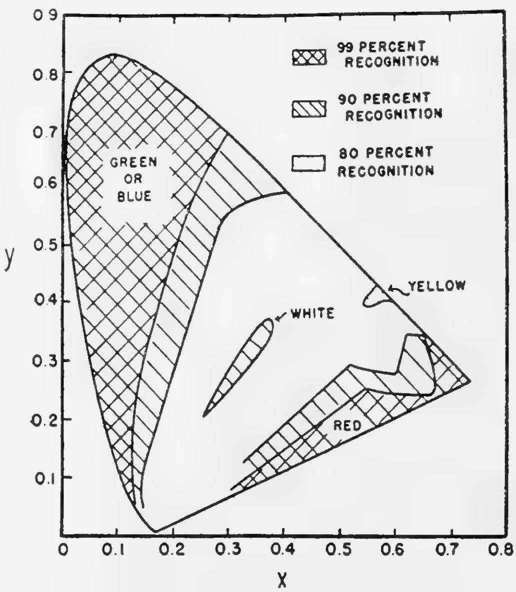

Fig. 29. The most recognizable colors for signalling at night. This plot employs the I.C.I. chromaticity diagram which shows the location of individual colors in a color triangle in terms of the percentage of red (x) and green (y) needed in a matching red-green-blue mixture. For details, see Appendix description. (After Hill, **40**)

Fig. 29. The two intensities of light gave pretty much the same result and Fig. 29 contains the data for only the two mile-candle light. The double-hatched areas are those colors which could be recognized correctly 99 percent of the time. The single-hatched areas show those colors which could be recognized correctly 90 percent of the time. There was no yellow or orange which could be correctly recognized 90 percent of the time, but the chart shows those yellows which could be recognized correctly 80 percent of the time.

A study of this magnitude is an enormous undertaking, and the importance of these results must not be minimized. But it must be recognized that this is far from a complete investigation. The parameters of luminous intensity, time of exposure, background luminance, and number of colored lights need to be explored more thoroughly before we shall have the complete story.

Size and Visibility. As in the case of visual acuity, size is an important factor in visibility. There are two ways in which the relationship may be illustrated. Fig. 30 shows the illuminance on the eye required

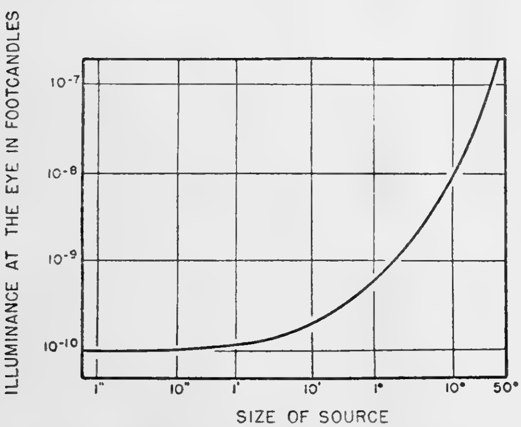

Fig. 30. Threshold visibility for light sources of various sizes in terms of illuminance produced at the eye. (From data compiled by Bouma, **5**)

for threshold visibility of light sources of various sizes. These data are average data for six experiments on this function. This figure makes it clear that point sources of light are more efficient than large areas.

The other way of showing data of this sort is in terms of the luminance of the object. Fig. 31 shows two sets of data taken from experiments by Lash and Prideaux (**51**) and Blackwell (**2**). The two sets of data do not agree very well but they show roughly the same function: large areas are much more visible than small areas of equal luminance. Offhand, this might seem to contradict the findings of Bouma cited above. The apparent discrepancy can be resolved, however, with a little thought about the kinds of measurements these represent.

Bouma's measurements on the illuminance on the eye represent a certain number of lumens deposited on the area of the pupillary opening. The data of Blackwell, on the other hand, are in terms of luminance, i.e., the number of lumens emitted per unit area of the target. If the lumens emitted *per unit area*, i.e., the luminance, is held constant, the target with the larger area will obviously emit the greater *total* number of lumens, and so will deposit the greater number of lumens per unit area on the eye. In short, when the appropriate conversions

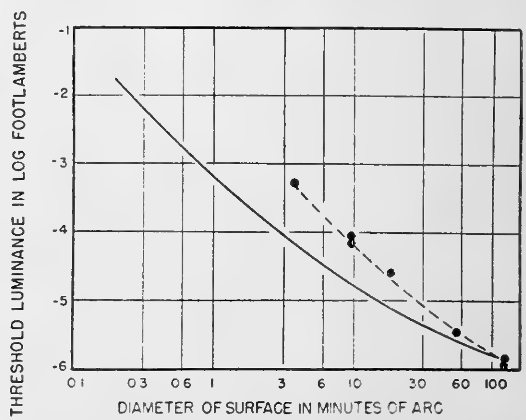

Fig. 31. The visibility of light sources at night in terms of the luminance of the source. (After Lash and Prideaux, **51**, and Blackwell, **2**)

are made, the data of Fig. 30 agree well with those in Fig. 31.

Time and Visibility. In considering how time relationships affect visibility, we have to consider two cases: (a) when the observer knows where the light source is and looks in that direction, and (b) when he has to hunt for it. In both cases a steady light source can be seen at much lower intensities than a flashing light source.

For the first case, i.e., when the location of the source is known, the intensity of a just visible flash of light bears this relation to that of a steady light:

$$E = E_0 \left(\frac{t + a}{t} \right),$$

where E = the illuminance at the eye produced by the just visible flashing light,

E_o = illuminance of the just visible steady light,

t = duration of the flash (in seconds), and

a = a constant equal to 0.21 seconds.

This function has also been plotted in Fig. 32 to show its nature. These data are from a study by Blondel and Rey (**4**) and are average data for 17 observers. Note that a flash of light which lasts about a half-second or less must be much more intense than a steady light in order to be seen. A flash of light which lasts a half-second or more is almost as visible as a steady light of the same intensity.

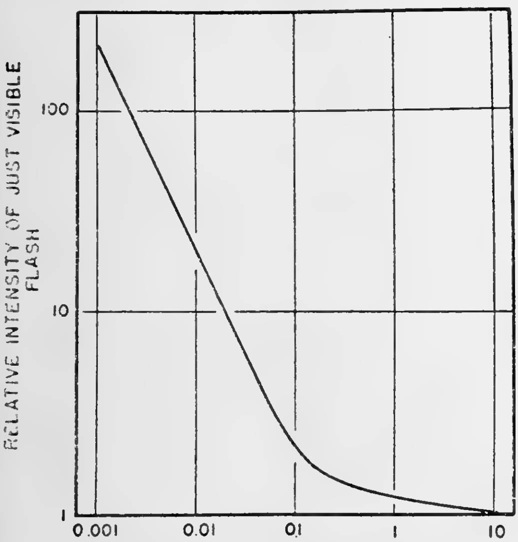

Fig. 32. Intensity of just visible flashes of light as a function of the duration of the flash. (After Blondel and Rey, **4**)

For the second case, i.e., where the location of the light is not known and the observer must hunt for it, the situation is much worse. Although the IES Lighting Handbook (**99**) gives an equation for this function, it is questionable whether any function can be trusted to correctly represent the facts. Some of the factors which must be considered are the length of time spent in hunting, the area searched, the area on the retina which is sensitive enough to see a light of that intensity, the kinds of eye movements the observer makes, how systematic a searcher he is, and so on. Only one thing can be stated with certainty. The values in Fig. 32 are minimum values. They show how intensity is related to time under the best seeing conditions. If the observer is required to search for a light, a large safety factor must be utilized.

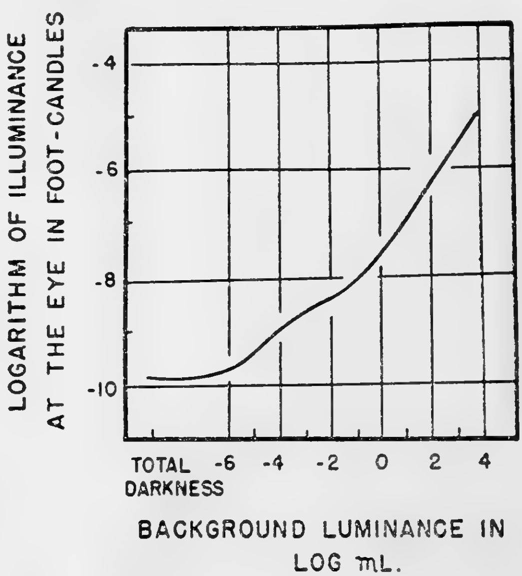

Fig. 33. Visibility of point sources as a function of background luminance. (After Knoll *et al.*, **48**)

Background Luminance. The visibility of a point source is a function of the background luminance. The data in Fig. 33 on this function are taken from the study by Knoll *et al.* (**48**). At starlight levels, a point source can be seen if it is very dim. As the background luminance increases, the threshold intensity increases rapidly.

Luminance Contrast. Our best information about the visibility of non-luminous areas comes from an extensive study carried out by Blackwell (**2**) during the war. Spots of light were projected onto a screen 60 feet away from a group of observers who reported whether they had seen the spot. A large number of such presentations were made with varying luminances, spots of various sizes, and with background luminances varying from full daylight to slightly less than starlight levels. In all, more than *two million* observations were recorded. Some 450,000 of these have now been analyzed, and they constitute what is probably the largest single study of human vision which has been reported to date.

There are so many ways in which the data of this study can be exhibited that it is necessary to select a few charts from it. Fig. 34 shows the least perceptible luminance contrast which can be seen by the normal eye against various background luminances.

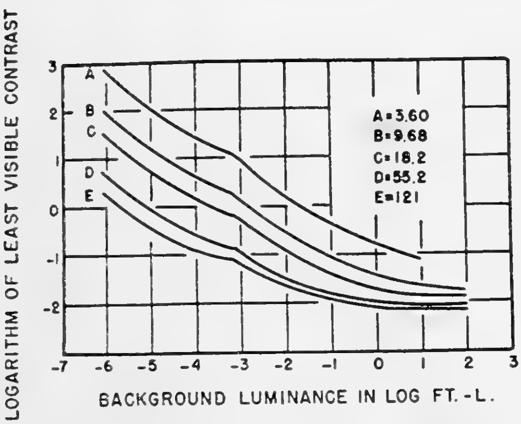

Fig. 34. Visibility of objects as a function of background luminance and size of object. The object sizes, A, B, C, D, and E, are in minutes of visual angle. (After Blackwell, **2**)

The different curves represent the data for test objects of various sizes. Two relationships are shown very clearly here: (a) as luminance decreases, the luminance contrast of the just perceptible object increases, and (b) at any luminance level, small objects have to have more contrast to be seen. The latter relationship is demonstrated more clearly in Fig. 35. There the size of the object is plotted against the least perceptible luminance contrast. In this case, the different curves represent the data obtained at the various luminance levels.

Applications of Contrast Data. The whole art of camouflage depends on methods whereby low color and brightness contrast are used to conceal objects by decreasing their visibility. Standard camouflage for naval aircraft consists in painting the underside a light color so as to present low contrast against the sky when seen from

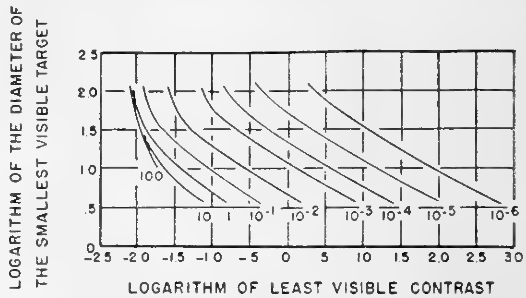

Fig. 35. These are the same data as in Fig. 34, replotted in another way. The different curves are for various luminance levels in footlamberts (After Blackwell, **2**)

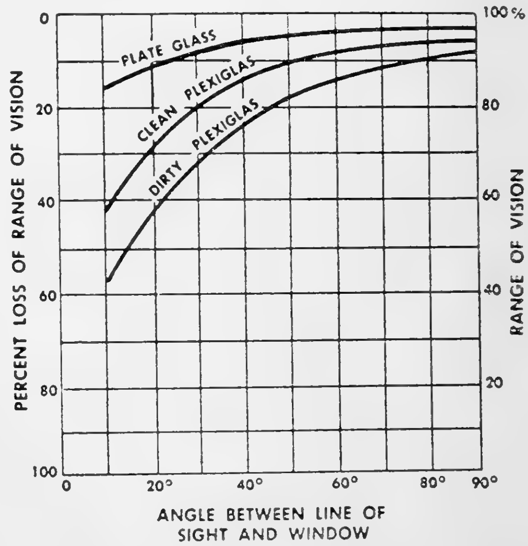

Fig. 36. The importance of luminance contrast in a practical situation. (After Olenski and Goodden, **71**)

below, and the top side blue to match the sea when seen from above. When aircraft operate predominantly in one type of combat environment, as do naval aircraft, satisfactory camouflage can be achieved easily. If, on the other hand, the aircraft must operate under a wide variety of environmental conditions, a camouflage for all of these conditions is difficult if not impossible to achieve.

Another practical illustration of the importance of brightness contrast is shown in Fig. 36, which scarcely needs any interpretation. These data were obtained by subjects who looked through aircraft glass, clean plexiglas, and dirty plexiglas. This figure shows that dust and grease on windshields act as an effective screen between the pilot and the outside world. Particles of dirt and grease scatter light haphazardly into the bundle of light rays which form an image of the object on the retina. This decreases the contrast and destroys the sharpness of the image. Diminished contrast also results from scratching or fogging of the transparent material, because each scratch or water droplet is a source of scattered light.

Summary

In this chapter we have tried to introduce some basic concepts, units of measurement, and functions in the field of visual psychophysiology. Although this is a long chapter, it is short in relation to the amount of research it has attempted to summarize. It has obviously been necessary to select from among the possible topics we might have covered. This is by no means the complete story of how we see, but it should provide some understanding of the kinds of data which can be put to use in applied problems.

Our second aim in this chapter has been to point out research needs in this field. We saw that much visual research needs reinvestigation because (a) it has not been sufficiently thorough to be of use to the applied visual scientist, and (b) the subjects have usually been so small in number—often only one—that the applied visual scientist cannot be sure whether the functions are representative of the average eye. The thorough reinvestigation of many of these functions with many subjects constitutes an enormous undertaking, but the practical benefits to be derived from the data are great.

Appendix

Measurement and Nomenclature in Visual Science

The Concepts of Light and Color

One difficulty with the study of light and color is that it does not fall exclusively within any of the major domains of science. It is now generally understood even by lay people that we see things either because they emit radiant energy or because radiant energy is reflected from them. In order for any object to be seen, radiant energy must come from the object and get into the eye. Radiant energy is similar to any other form of energy and its measurement is strictly a physical problem. But in order to be "seen," radiant energy must be converted into nervous energy by means of photochemical transformations in the eye and this nervous energy must then be transferred to the brain via the optic nerve. Both of these processes fall within the purview of the physiologist. Finally, the nervous energy in the brain must somehow be transformed into the inner experiences, sensations and perceptions, which constitute the "stuff" of seeing. The study of this process falls into the province of the psychologist. It is difficult, of course, to study this chain of events in its entirety, and one source of difficulty in the past has been that each group of scientists has had its own terminology and frame of reference for studying the same phenomena. A synthesis of these points of view is essential to the proper understanding of visual data.

Light and Color. When an observer's eye is stimulated by radiant energy and he reports that he sees light, the "light" he is referring to is not the same as radiant energy. It is radiant energy which has somehow been transformed into a part of a person's experience. Two definitions of light which have been adopted as standard are the following:

> For purposes of illuminating engineering, *light* is radiant energy . . . evaluated according to its capacity to produce visual sensation. (**100**)
>
> Light is that aspect of radiant energy of which a human observer is aware through the visual sensations which arise from the stimulation of the retina of the eye. (**18**)

Color is almost synonymous with light and has been defined as follows:

> Color consists of the characteristics of

light other than spatial and temporal inhomogeneities. (**18**)

The similarity between the definitions for light and color means, in more familiar terms, that white, gray, and black are colors. To distinguish them from what most people mean by color, the experiences of white, gray, and black are sometimes called achromatic colors; the ordinary colors like green, blue, and red are called chromatic colors. The characteristic thing about all

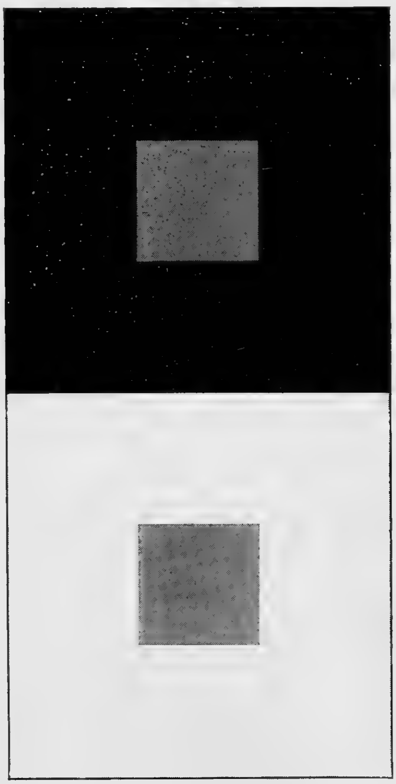

Fig. 37. This illustrates the phenomenon of simultaneous brightness contrast. The two gray areas are equal in reflectance. Do they look equally bright?

these definitions is that they involve sensations and thus enter into the province of psychology.

In most texts on vision and in the Illuminating Engineering Nomenclature and Photometric Standards approved in 1942 by the American Standards Association (**100**), there are two sets of terms and units to cover these two aspects of visual stimuli: the purely physical aspect (radiant energy), and the psychophysical aspect (radiant energy as evaluated by a conscious human observer, i.e., light and color). This is essentially the distinction which was made and defined in 1922 by the Committee on Colorimetry of the Optical Society of America (**89**) and in 1931 by the International Commission on Illumination. A considerable amount of visual research since then has shown that reports prepared by both groups required modification, and in 1933, the Optical Society of America appointed another Committee on Colorimetry (hereafter referred to as the OSA Committee) to consider revision and expansion of these earlier reports. Some of the differences of opinion among the various committee members are summarized in an article by Jones (**42**). It may be sufficient to point out that it required ten years of work by this committee before unanimous agreement could be reached.

Sensation. The concepts of light and color as defined above have been termed psychophysical concepts by the OSA Committee (**16**) because they cannot be identified simply with radiant energy nor with visual sensation. Both are operationally defined, i.e., the procedures used for evaluating light and color constitute the definitions of these concepts. But the OSA Committee recognized what psychologists have long known: there are differences between the magnitudes which measure light and the magnitudes of our inner experiences, i.e., sensations. The distinction can be shown by a simple experiment: Cover one eye tightly with a patch so that no light can reach it. Wear this patch in ordinary illuminations for a half hour. Now go into a dark room and look at a very dim bulb, first with the eye that has been covered, then with the other eye. The light from the bulb is unchanged, yet the visual sensations—the subjective impressions of brightness—are considerably different to the two eyes.

Another simple demonstration is the one shown in Fig. 37. The gray areas in the centers of both squares have the same amount of light coming from them (assuming that the page is held so that it is evenly illuminated), yet most people will agree that the gray area which is surrounded by the black looks brighter than the other gray area. Psychologists call this phenomenon simultaneous brightness contrast. Its importance here is that it points up a fundamental difference between amount of light, operationally defined, and the magnitude of the resulting sensation. The amounts of light from the two areas are the same (as can be verified by looking at both gray areas through holes in a piece of cardboard so that only the gray areas are visible) but the amounts of the sensations are different.

Considerations such as these have forced the OSA Committee to the position that the psychophysical concepts of light and color must be differentiated from those which refer directly to our sensations. Procedures for measuring the magnitudes of our sensations have been developed by psychologists, and the sensation magnitudes obtained are not the same as those which result from the measurement of light and color.

Comparable Concepts in Physics, Psychophysics, and Psychology. It is instructive and helpful to list parallel concepts which have been suggested by the OSA Committee, and Table II has been adapted from the committee's report. This is not a complete listing of concepts, since there are a number of others for measuring radiant flux and luminous flux which are not given here (see Table III, however). Three important points about this table are (a) that there are some blank places in it, i.e., there are no terms for certain attributes of our sensations to match characteristics of radiant energy and light; (b) certain new terms have been introduced, e.g., chromaticness, luminance, illuminance; and (c) the term brightness has been reserved to refer to that attribute of sensation which corresponds most closely to the physical and psychophysical concepts of radiant and luminous flux.

TABLE II

Some Parallel Concepts Relating to Light and Color

Adapted from a Report by the Committee on Colorimetry of the Optical Society of America (**16**)

Physics	Psychophysics	Psychology
Visual stimulus	Light	Visual sensation
Radiant energy Spectral composition	Luminous energy Color	Color sensation
Characteristics of radiant energy:	Characteristics of light (=color):	Attributes of color sensation:
A. Radiant flux	A. Luminous flux	A. Brightness
1. Radiance	1. Luminance	
2. Irradiance	2. Illuminance	
B. Spectral energy distribution	B. Chromaticity	B. Chromaticness
1. Relative spectral composition	1. Dominant wave-length	1. Hue
2. Radiant purity	2. Colorimetric purity	2. Saturation

Perception. There is still another class of concepts which, the OSA Committee recognizes, relate to our experiences of light and color. These are the perceptual concepts. They, like the concepts relating to sensation, fall entirely within the province of psychology. Psychologists are not entirely agreed on the differences between sensations and perceptions, but in a general kind of way the two terms are useful in distinguishing between the relatively simple, as compared with the relatively more complex, attributes of our inner experiences. An example might make the distinction somewhat more clear. It is possible to get a piece of blue paper which will match the blue of the sky in color. The blueness of both the sky and paper is a relatively simple aspect of our inner experiences and is called a sensation. But the sky does not look like the paper. The sky has a filmy, rather transparent appearance, while the color of the paper appears to reside in a surface. These "modes of appearance" are more complex than our sensations and they have been called perceptual attributes. Glossiness, lustre, transparency, sparkle, etc., are perceptual concepts. They have not been listed in Table II because they do not, in general, parallel the other three types of concepts. The perceptual concepts are the least well understood of all those we have discussed here. At present, psychologists have done little more than to classify some of these perceptual elements and to study some of the conditions which are necessary for their appearance. Little quantitative work has been done in this area.

With this as a background, we may now proceed to examine the kinds of measurements which may be made on visual stimuli.

Radiometry

Radiometry is the measurement of radiant energy in purely physical terms. The instruments, techniques, and units of measurement in radiometry have all been carefully standardized, and there is good agreement among physical scientists about them. They are difficult measurements to make, however, and this probably accounts for the fact that so few visual studies have measured stimuli in these terms. An excellent summary of the concepts, methods and units of radiometry is contained in a report of the OSA Committee (**17**).

Radiant Energy. Radiant energy, like all other forms of energy, may be measured and specified quantitatively in terms of the erg, the c.g.s. (centimeter, gram, second) unit of energy. Since the erg is a very small unit, the m.k.s. (meter, kilogram, second) unit, the joule (equal to 10,000,000 ergs), is frequently used instead.

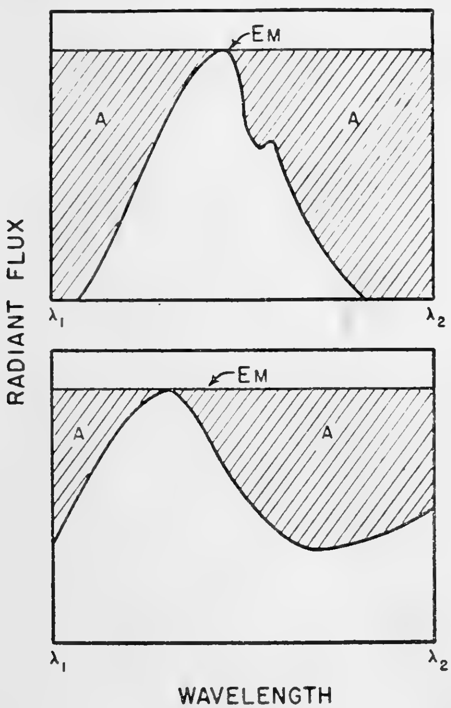

Fig. 38. To be meaningful in visual work, the specification of visual stimuli in physical terms must show the radiant flux per unit wavelength interval as is illustrated here. These diagrams also illustrate the concept of radiant purity. The upper distribution of radiant energy exhibits more purity than the lower one.

The evaluation of radiant energy as such, either in terms of ergs or joules, is appropriate only for the specification of the total energy radiated or contained in a given volume of space (radiant density), or received in any specific instance. This is not an appropriate measure to use in visual work because the eye is not an accumulator—it does not respond to the total amount of energy received on the retina over a period of time, but rather to the rate (or density with respect to time) at which the radiant energy is deposited on the retina.

Radiant Flux. The rate of transfer of radiant energy is termed radiant power, or, more commonly, radiant flux. It may be measured in c.g.s. units, ergs per second, or in m.k.s. units, watts (joules per second), and may be measured for all the wavelengths in a bundle of radiant energy, or for each wavelength interval separately.

Radiant Intensity. The radiant flux radiated within a unit solid angle (steradian) around a point source defines the radiant intensity of the source. Radiant intensity is measured in ergs per second per unit solid angle (ergs/sec. $\times$ ω) or watts per unit solid angle (watts $\times$ ω). If the source is not a point source, measurements must be made at a distance much greater than the largest dimension of the source, otherwise the measurement of solid angles around the source becomes meaningless.

Radiant Emittance. The intensity of a source may also be specified in terms of the total radiant flux emitted in all directions per unit area of the source, radiant emittance. Radiant emittance may be expressed in ergs per second per square centimeter (ergs/sec. $\times$ cm.^{2}), or watts per square meter (watts/m.^{2}).

Irradiance. The amount of radiant flux falling on a surface is termed irradiance, and may be measured in ergs per second per square centimeter (ergs/sec. $\times$ cm.^{2}), or watts per square meter (watts/m.^{2}). If the source is small enough to be a point source, or if the distances involved are large enough so that the source may be considered a point source, the irradiance is proportional to the intensity of the source and inversely proportional to the distance between the source and the surface.

Radiance. Because many sources of radiant energy cannot be considered point sources, it is inappropriate to specify them in terms of radiant intensity. The term radiance is used to cover large sources. In contrast to radiant emittance, radiance is a directional quantity; it measures radiant flux emitted in a certain direction. Radiance is measured in ergs per second per unit solid angle per square centimeter (ergs/sec. $\times$ ω $\times$ cm.^{2}) or watts per unit solid angle per meter (watts/ω $\times$ m.^{2}). This is a measure of the radiant flux emitted in a solid angle centered on the surface, or on the projected area of the surface, if it is not viewed perpendicularly. Since radiance applies to large sources or surfaces, it is usually measured by selecting an area of the source small enough, and a measuring distance great enough, so that the measurement becomes one of radiant intensity.

As we have already seen, the eye does not respond to all wavelengths of radiant energy. If visual stimuli are specified in physical terms, the specifications must be in terms of selected, narrow wavelength intervals if they are to be meaningful. Two such spectral distribution curves are shown in Fig. 38. They are important for exhibiting the quality of the radiant energy in a visual stimulus.

Radiant Purity. Fig. 38 is also convenient for defining the concept of radiant purity. Radiant purity between the wavelengths λ_1 and λ_2 is the ratio of the shaded areas marked A to the entire area under the line E_m. The higher the ratio, the greater is the radiant purity. In Fig. 38, for example, the distribution of radiant flux in the upper figure is more pure than the distribution in the lower figure. This is a way of saying that a greater proportion of the radiant flux is concentrated in a relatively narrow part of the spectrum.

Other Concepts in Radiometry. Other concepts in radiometry, together with their defining equations, are listed in Table III.

TABLE III

Comparable Units and Measurements in Radiometry and Photometry

Adapted from a Report of the Committee on Colorimetry of the Optical Society of America (**18**)

Physical				Psychophysical			
Radiator (source of radiant energy) Radiation (process)				Luminator (source of luminous energy) Lumination (process)			
Radiometry	Symbol	c.g.s. unit	m.k.s. unit	Photometry	Symbol	c.g.s. unit	m.k.s. unit
Radiant energy	U	erg	joule	Luminous energy	Q	lumerg	talbot
Radiant density	u	erg/cm.^{3}	joule/m.^{3}	Luminous density	q	lumerg/cm.^{3}	talbot/m.^{3}
Radiant flux	P	erg/sec.	watt	Luminous flux	F	lumerg/sec.	lumen
Radiant emittance	W	erg/sec. × cm.^{2}	watt/m.^{2}	Luminous emittance	L	lumerg/sec. × cm.^{2}	lumen/m.^{2}
Radiant intensity	J	erg/sec. × ω	watt/ω	Luminous intensity	I	lumerg/sec. × ω	lumen/ω(candle)
Radiance	N	erg/sec. × ω × cm.^{2}	watt/ω × m.^{2}	Luminance	B	lumerg/sec. × ω × cm.^{2}	lumen/ω × m.^{2} (candle/m.^{2})
Irradiance	H	erg/sec. × cm.^{2}	watt/m.^{2}	Illuminance	E	lumerg/sec. × cm.^{2}	lumen/m.^{2}(lux)

Photometry

Corresponding to each of the radiometric units there is a photometric unit for measuring the amount of light. The names, units, and defining equations recommended by the OSA Committee are shown in Table III. This table also shows the parallelism between the two kinds of units. Every photometric unit in this table can be defined in terms of its corresponding radiometric unit simply by substituting the word "luminous" for "radiant." The word luminous in this context means radiant energy evaluated according to the sensation produced in an average eye. As has already been indicated, this sytem of units and nomenclature differs from the one standardized by the American Standards Association in 1942 (**100**). Since the latter defines the units which most people are familiar with, it may be worthwhile to review them briefly.

Luminous Intensity. The unit of luminous intensity or candle-power is the International Candle, an arbitrary intensity agreed upon and maintained by the three national standardizing laboratories of France, Great Britain and the United States. This concept is unchanged in the new nomenclature.

Another unit of luminous intensity is the Hefner candle, equal approximately to 0.90 of an international candle. Although this unit is commonly encountered in visual literature, particularly German literature, published before 1939, it is seldom used now.

Luminous Flux. The old unit of luminous flux, the lumen, was also adopted without change by the OSA Committee. A lumen is equal to the luminous flux emitted in a unit solid angle by a point source of one international candle.

Illumination. Illumination (illuminance in the new OSA nomenclature) is the density of the luminous flux deposited on a surface. The lux, or meter-candle, is a convenient unit of illumination. It is the illumination on a surface one square meter in area when this surface is receiving a uniformly distributed flux of one lumen. By virtue of the definition of a lumen above, a lux is also the illumination deposited on the inner surface of a sphere having a radius of one meter with a point source of one international candle at the center.

Several other units of illumination are used in visual work and Table IV shows how they are related to each other. The phot and foot-candle, for example, correspond to the lux with the words "centimeter" and "foot" substituted for "meter" in their respective definitions.

Brightness. In the older terminology, brightness was a confusing concept, not only because of the diversity of units used to measure it (see Table V), but also because the concept implied two different types of measurements: one in terms of luminous intensity per unit area of the surface (luminance in the newer terminology); the other in terms of the luminous flux emitted by or reflected from the surface (luminous emittance in the newer terminology). In Table V, units in candles per unit area (mm., cm., etc.) are of the first type; units in lamberts (or any lambert combination) are of the second type (**75**). In the case of perfectly diffusing surfaces, i.e., those square centimeter. The luminous emittance of a surface is the total luminous flux emitted in all directions per unit area of the surface. The unit is the lumen per square centimeter. For a perfect diffuser, the luminous emittance (in lumens/cm.^{2}) is π times the luminance in candles/cm.^{2} For such a surface, the luminous emittance (in lumens/cm.^{2}) is also exactly equal to the luminance in lamberts.

Brilliance. In the report of the first OSA Committee on Colorimetry, Troland (**89**) suggested the term "brilliance" to refer to the magnitude of the visual sensation as distinct from its corresponding photometric quantity, brightness. The term brilliance was never generally adopted by psychologists even though one might have anticipated that they would have welcomed it.

TABLE IV

CONVERSION FACTORS FOR ILLUMINATION (ILLUMINANCE) UNITS

(Value in unit in left hand column times the conversion factor equals the value in unit shown at the top of the column.)

	Foot-candle	Mile-candle	Sea mile-candle	Centimeter-candle	Milliphot	Meter-candle	Kilometer-candle
Foot-candle	1	2.788×10^{7}	3.697×10^{7}	1.076×10^{-3}	1.076	1.076×10	1.076×10^{7}
Mile-candle	3.587×10^{-8}	1	1.326	3.863×10^{-11}	3.863×10^{-8}	3.863×10^{-7}	3.863×10^{-1}
Sea mile-candle	2.705×10^{-8}	7.540×10^{-1}	1	2.911×10^{-11}	2.911×10^{-8}	2.911×10^{-7}	2.911×10^{-1}
Centimeter-candle (Phot)	9.290×10^{2}	2.589×10^{10}	3.435×10^{10}	1	1×10^{3}	1×10^{4}	1×10^{10}
Milliphot	9.290×10^{-1}	2.589×10^{7}	3.435×10^{7}	1×10^{-3}	1	1×10	1×10^{7}
Meter-candle (Lux)	9.290×10^{-2}	2.589×10^{6}	3.435×10^{6}	1×10^{-4}	1×10^{-1}	1	1×10^{6}
Kilometer-candle	9.290×10^{-8}	2.589	3.435	1×10^{-10}	1×10^{-7}	1×10^{-6}	1

which follow what is called the cosine law, conversions between the two kinds of measurements are simple and straightforward. Actually, very few surfaces are perfect diffusers and many do not even approach this condition closely. In practice, however, visual scientists have ignored any implied difference between the two kinds of units and have used them interchangeably by means of the conversion factors shown in Table V. Few tables of this kind have ever been published, and this one should be useful in helping the reader to convert the measurements of other visual experiments into those used here. It should also impress him with the unnecessary duplication of units which has been common.

It is clear now why the OSA Committee has suggested two new terms, luminance and luminous emittance, to replace brightness in the old sense. Luminance is the luminous flux per unit solid angle emitted per unit projected area of the source. Although all the units in Table V can be considered measures of luminance, the OSA Committee recommends one, the candle per

Still Other Units. Other photometric units have been suggested by Moon (**67**) for international adoption. His units include pharos (for luminous flux), pharosage (for illuminance), helios (for luminance), etc. These terms are so unlike those which have been in common use that they will not be discussed here. In England the units glim, scot and nox have been proposed for low-luminance measurements (**102**). These are equal to 10^{-3} foot lamberts, 10^{-5} equivalent meter candles or 1 milli-microlambert, and 10^{-5} meter-candles, respectively.

Standardization Needed. The foregoing discussion has undoubtedly impressed upon the reader the need for standardization and simplification of photometric concepts and measurements. Lest the reader get the impression that this is, however, an internal affair and of no great concern to Naval scientists and engineers, the writer would like to express his opinion that the problem of standardization is one of the most pressing problems confronting visual scientists at the present time. As a practical illustration of its importance,

TABLE V

Conversion Factors for Brightness (Luminance and Luminous Emittance) Units

(Value in Unit in Left Hand Column Times the Conversion Factor Equals the Value in Unit Shown at the Top of the Column.)

	C/mm^2	C/cm^2	C/m^2	Stilb (H)	C/in^2	C/ft^2	L	mL	μL	mμL	μμL	Apostilb (H)	Ft.-L	Photons*
Candles per Square Millimeter	1	1×10^{2}	1×10^{6}	1.111×10^{2}	6.452×10^{2}	9.290×10^{4}	3.142×10^{2}	3.142×10^{5}	3.142×10^{8}	3.142×10^{11}	3.142×10^{14}	3.491×10^{6}	2.919×10^{5}	7.854×10^{5}
Candles per Square Centimeter (C.I.E. Stilb)	1×10^{-2}	1	1×10^{4}	1.111	6.452	9.290×10^{2}	3.142	3.142×10^{3}	3.142×10^{6}	3.142×10^{9}	3.142×10^{12}	3.491×10^{4}	2.919×10^{3}	7.854×10^{3}
Candles per Square Meter	1×10^{-6}	1×10^{-4}	1	1.111×10^{-4}	6.452×10^{-4}	9.290×10^{-2}	3.142×10^{-4}	3.142×10^{-1}	3.142×10^{2}	3.142×10^{5}	3.142×10^{8}	3.491	2.919×10^{-1}	7.854×10^{-1}
Hefner Candles per Square Centimeter (Stilb (H))	9.0×10^{-3}	9.0×10^{-1}	9.0×10^{3}	1	5.806	8.361×10^{2}	2.828	2.828×10^{3}	2.828×10^{6}	2.828×10^{9}	2.828×10^{12}	3.142×10^{4}	2.627×10^{3}	7.069×10^{3}
Candles per Square Inch	1.550×10^{-3}	1.550×10^{-1}	1.550×10^{3}	1.722×10^{-1}	1	1.440×10^{2}	4.869×10^{-1}	4.869×10^{2}	4.869×10^{5}	4.869×10^{8}	4.869×10^{11}	5.411×10^{3}	4.524×10^{2}	1.217×10^{3}
Candles per Square Foot	1.076×10^{-5}	1.076×10^{-3}	1.076×10	1.196×10^{-3}	6.944×10^{-3}	1	3.382×10^{-3}	3.382	3.382×10^{3}	3.382×10^{6}	3.382×10^{9}	3.758×10	3.142	8.454
Lamberts (Equivalent Centimeter Candles, Apparent Lumens per Square Centimeter)	3.183×10^{-3}	3.183×10^{-1}	3.183×10^{3}	3.537×10^{-1}	2.054	2.957×10^{2}	1	1×10^{3}	1×10^{6}	1×10^{9}	1×10^{12}	1.111×10^{4}	9.290×10^{2}	2.500×10^{3}
Millilamberts	3.183×10^{-6}	3.183×10^{-4}	3.183	3.537×10^{-4}	2.054×10^{-3}	2.957×10^{-1}	1×10^{-3}	1	1×10^{3}	1×10^{6}	1×10^{9}	1.111×10	9.290×10^{-1}	2.500
Microlamberts	3.183×10^{-9}	3.183×10^{-7}	3.183×10^{-3}	3.537×10^{-7}	2.054×10^{-6}	2.957×10^{-4}	1×10^{-6}	1×10^{-3}	1	1×10^{3}	1×10^{6}	1.111×10^{-2}	9.290×10^{-4}	2.500×10^{-3}
Milli-Microlamberts (Micro-Millilamberts)	3.183×10^{-12}	3.183×10^{-10}	3.183×10^{-6}	3.537×10^{-10}	2.054×10^{-9}	2.957×10^{-7}	1×10^{-9}	1×10^{-6}	1×10^{-3}	1	1×10^{3}	1.111×10^{-5}	9.290×10^{-7}	2.500×10^{-6}
Micro-Microlamberts	3.183×10^{-15}	3.183×10^{-13}	3.183×10^{-9}	3.537×10^{-13}	2.054×10^{-12}	2.957×10^{-10}	1×10^{-12}	1×10^{-9}	1×10^{-6}	1×10^{-3}	1	1.111×10^{-8}	9.290×10^{-10}	2.500×10^{-9}
Apostilb (Hefner Lumens per Square Foot)	2.864×10^{-7}	2.864×10^{-5}	2.864×10^{-1}	3.183×10^{-5}	1.848×10^{-4}	2.661×10^{-2}	9.0×10^{-5}	9.0×10^{-2}	9.0×10	9.0×10^{4}	9.0×10^{7}	1	8.360×10^{-2}	2.249
Foot-Lamberts (Equivalent Foot Candles; Apparent Foot Candles; Apparent Lumens per Square Foot)	3.426×10^{-6}	3.426×10^{-4}	3.426	3.807×10^{-4}	2.210×10^{-3}	3.183×10^{-1}	1.076×10^{-3}	1.076	1.076×10^{3}	1.076×10^{6}	1.076×10^{9}	1.196×10	1	2.691
Photons†	1.273×10^{-6}	1.273×10^{-4}	1.273	1.414×10^{-4}	8.213×10^{-4}	1.183×10^{-1}	4.000×10^{-4}	4.000×10^{-1}	4.000×10^{2}	4.000×10^{5}	4.000×10^{8}	4.444×10	3.716×10^{-1}	1

* In converting measures of brightness into photons, it is necessary to multiply the conversion factor by the square of the pupil diameter in millimeters.

† In converting photons to measures of brightness, it is necessary to divide the conversion factor by the square of the pupil diameter in millimeters.

during one publication year, 1944, there appeared in the *Journal of the Optical Society of America* 12 articles in which photometric measurements and concepts were used in connection with visual data. (This count does not include the reports of the OSA Committee.) Seven of these articles used the old terminology and concepts, two the new OSA nomenclature, and three the Moon (**67**) nomenclature. This situation is confusing enough to the visual expert when he tries to convert concepts and measurements of one kind to those of another. The Naval engineer or scientist, who may be unaware of this situation, however, is likely to conclude that an article which talks about "pharosage," "helios," and "heliosent," and expresses measures in "blondels," for example, has nothing in common with another article which talks about "illuminance" and "luminance," and expresses its measures in "millilamberts." The development of a common language is essential if the visual scientist wants to be understood not only by his colleagues, but by other scientists and engineers as well.

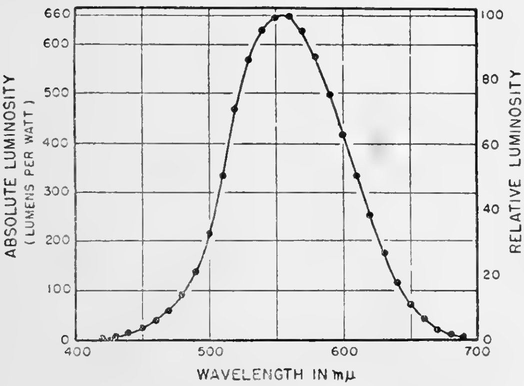

Fig. 39. Absolute and relative photopic luminosity values for radiant flux of various wavelengths computed for the average eye. (Data from I.C.I., **100**)

Standardization should be on an international scale and should be sponsored by the Bureau of Standards of the major countries of the world. Other government agencies, e.g., the Navy Department, can probably do much, however, to expedite the completion of such a program. In the meantime, individual investigators can also further this goal by using a consistent set of nomenclature and units in their own work. The proposals of the OSA Committee (**16, 17, 18, 19**) represent the considered opinions of a group of leading visual scientists in this country. They are consistent, simple and rational, and, as such, are recommended for use in all visual work.

The Techniques of Photometry. If the spectral energy distribution of a stimulus is known, it is possible to compute the corresponding photometric quantity without making any additional measurements. It is possible to do this because the spectral sensitivity of the "average" eye is known. Actually, of course, the "average" eye is an abstraction based on the data obtained from a fairly large sample of normal eyes. The average normal luminosity values adopted by the International Commission on Illumination and the American Standards Association (**100**) are shown in Fig. 39. This curve—the photopic relative luminosity curve discussed in the chapter above—shows that one watt of radiant flux of 555 mμ falling on an average eye will result in 660 lumens of light. One watt of radiant flux of 627 mμ will result in 200 lumens of light, etc. It is apparent, then, that if the radiant flux per wavelength interval is known for a stimulus, the corresponding amount of light seen by an average, light-adapted eye can be computed from the data in Fig. 39. Moon's textbook on illuminating engineering (**68**) actually defines photometric measures in this mathematical way. Thus, for example, he defines luminous flux as "radiant power evaluated with respect to the standard visibility [luminosity in the new OSA terminology] function."

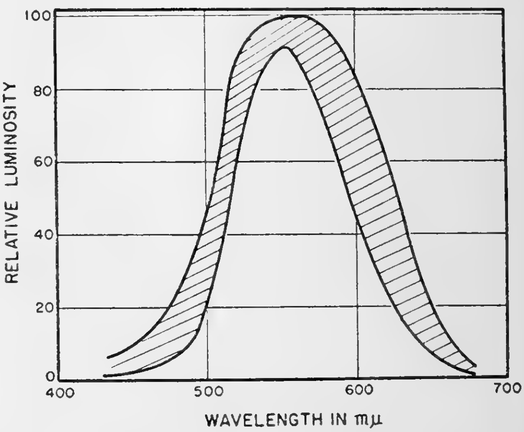

Fig. 40. The range of individual differences in the relative luminosity function among 52 observers studied by Gibson and Tyndall (**28**).

Actually, this method of photometry is very difficult and is practically never used by the ordinary visual scientist. Most commonly, as was described in the text above, the light to be evaluated is compared with a standard light of known luminosity using an observer's eye as a null indicator. A summary of these techniques is contained in a report of the OSA Committee (**18**) and in the comprehensive text by Walsh (**94**).

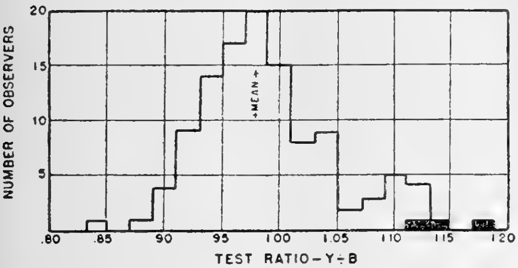

Fig. 41. Individual differences in photometric measurements made on a yellow and blue light by 114 observers. The solid areas are color blind observers. (Data from Crittenden and Richtmyer, **22**)

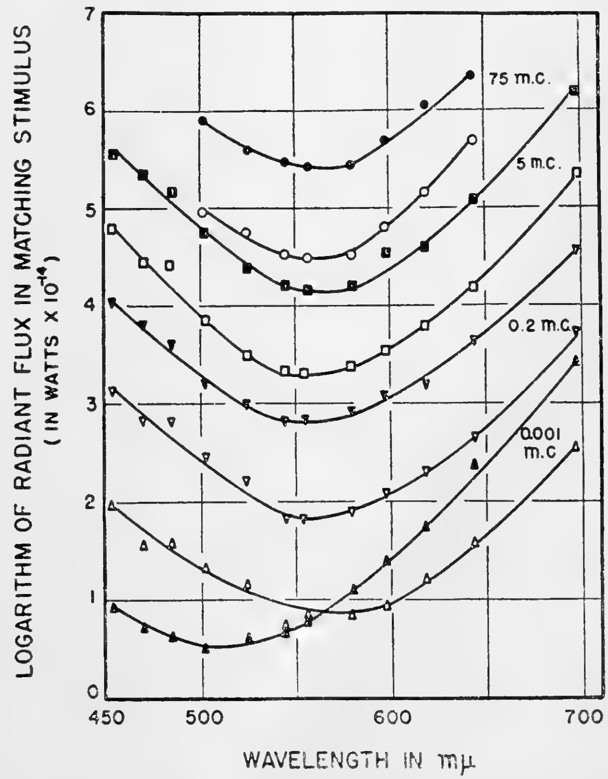

Fig. 42. The amount of radiant flux required by a completely dark-adapted eye to match four test luminances. The solid symbols are for a large test area (4° 49′); the open symbols for a small test area (57′). Note the complete change in the shape of one of the curves at the lowest luminance level. (Data from Sloan, **83**)

Individual Differences. Those techniques of photometry which make use of an observer's eye as a photosensitive indicator are simple and are perhaps the kinds most commonly used in visual work. Unfortunately, they introduce some special problems which most visual scientists prefer to ignore. One of these is the complication introduced by individual differences. In reading many reference works on vision it is easy to get the impression that the relative luminosity curve (Fig. 39) represents a precisely defined function for all people. Actually, individual differences in this function are enormous. Fig. 40 shows the range of luminosity values found among 52 observers studied by Gibson and Tyndall (**28**). Note that at 600 mμ, for example, the range of values is almost in the ratio of two to one. The actual percentages for the two most deviant observers are 44.9 and 81.9. Related data by Crittenden and Richtmyer (**22**) on 114 observers are shown in Fig. 41. Here the observers were required to measure photometrically standard yellow and blue lights. Note that some observers reported the yellow light to be roughly eight-tenths as intense as the blue light, while others reported the yellow light to be considerably more intense than the blue light. The important point for our purposes, and the point which is frequently not appreciated, is that differences in spectral sensitivity are large. This means that photometric values measured by different observers might vary greatly. Further, one cannot be sure that his own eye is representative of the "average" eye unless his eye has been carefully checked out against the standard luminosity function.

and 43. Fig. 42 shows two sets of data obtained by Sloan (**83**) with a completely dark-adapted eye and with two different sizes of test fields: 4° 49′ (solid points) and 57′ (open symbols). The observer looked at a test field split in the middle, one half of the field being filled with white light of the luminance levels shown in Fig. 42, the other filled with monochromatic spectral light (i.e., light composed of few wavelengths). The observer was required to adjust the intensity of the

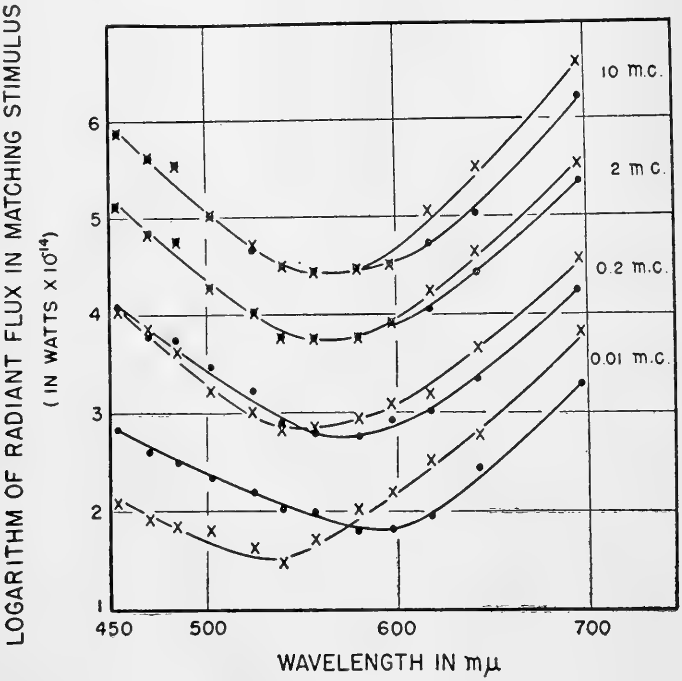

Fig. 43. The amount of radiant flux required by a light-adapted eye (crosses) and a dark-adapted eye (dots) to match four test luminances. The size of the test area is constant throughout. Note the differences between pairs of curves especially at the lowest luminance level. (Data from Sloan, **83**)

Low Luminance Photometry. On the very knotty and unsolved problem of low luminance photometry, we can do little more than discuss some aspects of the problem here. The interested reader may want to consult the more thorough discussions by Taylor (**86**) and Luckiesh and Taylor (**59**).

A great many visual functions are studied at luminances which approach the threshold of vision and are well below the cone threshold. In some cases the results of these studies show curious inconsistencies because the ordinary relationships between radiant energy and luminous energy do not hold. This is another way of saying that at luminances below the cone level the curve shown in Fig. 39 is no longer valid. This can be illustrated by the data in Figs. 42

monochromatic light until it just matched the white light.

Fig. 42 contains only a small portion of the data given by Sloan, but it illustrates an important point. When matches have to be made at fairly high luminance levels (above 0.2 meter candles), there appears to be a fairly constant difference between the curves for the large test area and those for the small test area. Most of this difference can be accounted for by the greater amount of radiant flux emitted from the larger field. If the data had been expressed in flux per unit area (radiant emittance), the curves would be nearly coincident. But notice what happens at the lowest matching luminance. The curves change completely. At the shorter wave-lengths, in fact, there is a complete reversal in the relationships:

more radiant flux is required to achieve a match with the smaller test patch, although in every other set of curves less radiant energy is required for the smaller test patch.

Fig. 43 shows still another state of affairs. Here matches were made with the same size of test field, 4° 49′, but the matches were made with the observer light-adapted (crosses) or dark-adapted (dots). Although the external conditions of the stimulus situation were exactly the same, entirely different psychophysical relationships were obtained with different states of adaptation of the eye.

The point of this discussion is that the ordinary photometric procedures which work well at high luminance levels do not work at very low luminance levels. The relationships between radiant flux and luminous flux vary with the size of the stimulus field, the state of adaptation of the eye and the area of the eye being stimulated. The problem became especially serious during the last war in connection with the photometry of luminescent and phosphorescent materials. Tentative procedures were worked out at the time in order that some degree of consistency could be achieved between different laboratories and between laboratories and manufacturers. The procedures were admittedly tentative and the problem of low luminance photometry still needs to be studied thoroughly. Examples of inconsistencies in visual data arising from this source have been pointed out in the chapter above.

Colorimetry

The measurement of the chromatic aspects of light is termed colorimetry. Here again this evaluation is made in terms of the sensation aroused in an average eye.

The Eye is an Integrator. The eye cannot resolve different wavelengths when they are combined in a ray of light. This is another way of saying that it is an integrating mechanism rather than an analyzing one. Fig. 44 shows five different combinations of wavelengths which look exactly the same to an average eye. The uppermost section, for example, shows a spectrum which contains equal amounts of all the visible wavelengths, i.e., an equal-energy spectrum. The second chart shows two single wavelengths which, when combined in the ratio of 1 to 0.78, produce exactly the same sensation as the equal-energy spectrum. The same is true for all the other combinations shown in Fig. 44. Even though the combination of wavelengths differs markedly in all cases, the color is exactly the same—very nearly white.

Color Mixtures. The integrating behavior of the eye provides the basis for the psychophysical specification of color. It was discovered a long time ago that every color can be matched by a mixture of three colored lights, red, green, and blue. This is very convenient because it means that any color can be specified in terms of how much of a standard red, green, and blue must be mixed together to match the unknown color. This information can then be charted so that any color can be compared with any other.

Chromaticity Diagram. The chart used for this purpose, Fig. 45, is called a chromaticity diagram. This one was standardized by the International Commission on Illumination in 1931. The red, green, and blue lights which were standardized are mathematically defined and are hypothetical colors. But the reader will not be far off if he regards them as being real colors.

Only two dimensions are needed on the chromaticity diagram because the amounts of green, blue, and red required to match a color are expressed in percentage terms, i.e., they add up to 100 percent. This means that if the amount of red and green are known, the amount of blue can be found by subtraction from one. It would take us far beyond the scope of this discussion to consider precisely how the position of a color can be computed on this chart. A summary of these techniques is contained in another report of the OSA Committee (**19**).

The location of the spectrum colors in the chromaticity diagram is shown by the roughly triangular curve in this figure. The numbers along the curve are wavelengths. Since the spectrum colors are the purest one can find, all other colors fall somewhere inside this boundary.

White. In a general kind of way, white results from a mixture of all wavelengths. It can be matched by very nearly equal proportions of red, green and blue and should be located somewhere near the center of this chart. Actually white is a hard color to define. Most people say that a Mazda lamp or fluorescent lamp gives off white light, although the lights are yellowish and bluish, respectively. Even the color of daylight varies considerably. It is more reddish in the morning and late afternoon, and its color depends greatly on the number of clouds in the air, the time of year, and so on. The International Commission on Illumination defined several whites and the one shown in Fig. 45 is Illuminant C. This is about the kind of light that comes from a summer sky when the air is clear, there are no clouds, and the sun is high. The point E, incidentally, is the color of all those different wavelength combinations shown in Fig. 44. It is a little more yellowish than Illuminant C.

Dominant Wavelength. In Fig. 45 are plotted the colors of a lemon (*L*) and ripe tomato (*T*). A line has been drawn from the point *W* through the color of the tomato, *T*, until it intersects the spectrum line at approximately 596 mμ. This point of intersection with the spectrum line defines the dominant wavelength of the color.

radiant purity are fairly well correlated—the greater the radiant purity of the stimulus, the greater the colorimetric purity of the light.

The Use of the ICI Diagram. A chromaticity diagram is very useful because it tells the visual scientist a lot about how the color looks. If the color is far out to the right, it looks reddish.

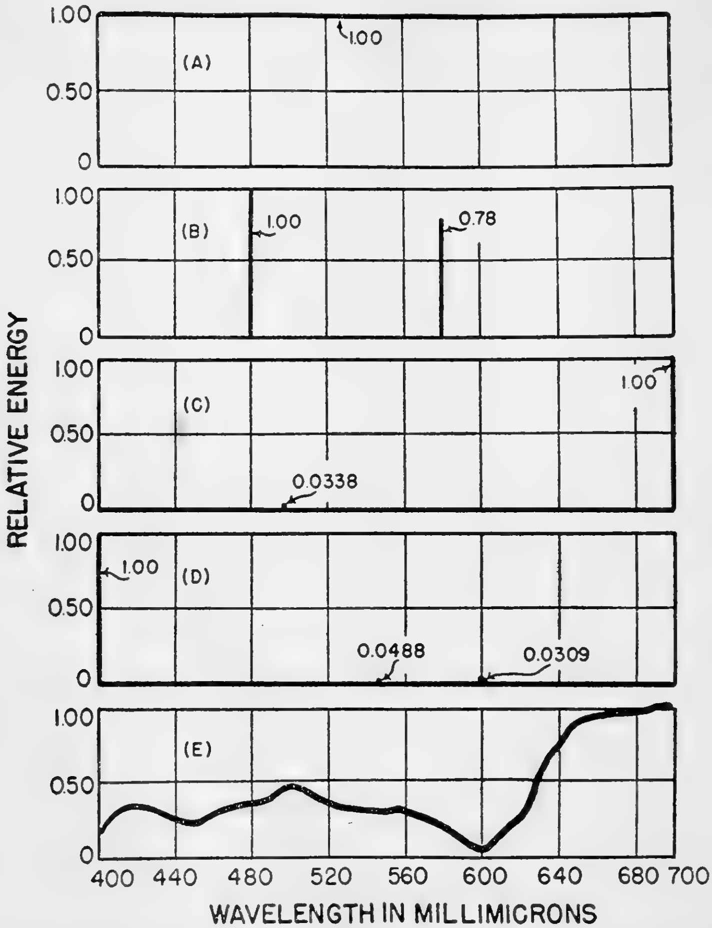

Fig. 44. All five combinations of wavelengths shown here look exactly alike to the eye—very nearly white. This illustrates the fact that the eye is an integrating mechanism rather than an analyzing one. (After Moon, **68**)

Colorimetric Purity. The distance of *T* along the line WS, expressed in percentage terms, defines the colorimetric purity of the color. The closer it is to *S*, the purer the color; the closer it is to *W*, the more "washed out" it looks. This is another way of saying that colorimetric purity is correlated with the attribute of our sensations called saturation. In general, colors with greater amounts of colorimetric purity look more saturated than colors with lesser amounts. As another general rule, colorimetric purity and

If the color is high on the chart, it looks greenish. If the color is near the lower left-hand corner, it does not contain much red or green and so looks bluish. If a color is near the straight line running from 380 mμ to 780 mμ, it will look purple. Purples do not occur in the spectrum; they are actually mixtures of red and blue.

Another advantage of the chromaticity diagram is that it immediately defines the mixtures which are obtained from any combination of colors. If 500 and 560 mμ are mixed, for example, the straight line joining these two points on the chromaticity diagram is the locus of all color combinations resulting from the mixtures.

Measurement of Sensations

The measurement of sensations is an interesting area of research which has hardly been investigated by psychologists. Many physicists doubt that sensations can be measured, but orderly and systematic work in this area is possible. As evidence of this, we may cite a very recent study on the relationship between luminance and brightness.

The Relationship Between Luminance and Brightness. In this study (**33**) each observer was required to adjust the luminance on one test area until it looked exactly half as bright as another. None of the observers expressed any doubts about being able to make these fractional judgments of brightness nor did the nature of the judgments seem unnatural or artificial to them. They were also able to make these judgments with considerable consistency.

The results of this investigation are shown in Fig. 46. The abscissa shows the luminance of the standard test area, the ordinate the luminance of the test area which looked half as bright. Three things are evident from this figure: (a) the results obtained with the different colors seem to agree with each other, (b) the agreement among the various points is almost as good as one gets with photometric measurements, and (c) a luminance which is half another luminance does not necessarily look half as bright.

The data cited above are preliminary, but they hold great promise for quantification in an area where it was once thought impossible. More work is under way to explore more fully still other parameters in connection with this function.

Chromaticity and Color Sensation. It is now known that equal distances on the ICI chromaticity diagram do not represent equally noticeable color differences. This is another way of saying that color sensation does not keep pace with chromaticity. Data on this point come from a study by MacAdam (**63**), who investigated chromaticity differences which could just be perceived by the eye. His data are shown in Fig. 47. The distances from the center of any ellipse to the boundary of the ellipse represent equally noticeable color differences. Actually, the ellipses in this figure are magnified 10 times.

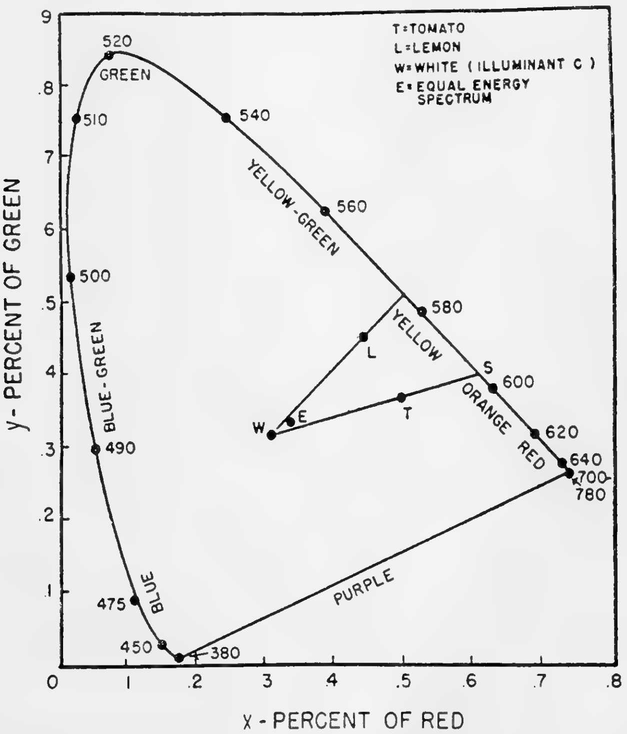

Fig. 45. The I.C.I. chromaticity diagram for plotting the locations of colors in terms of the percentage of standard red and green in red-green-blue mixtures required to match them.

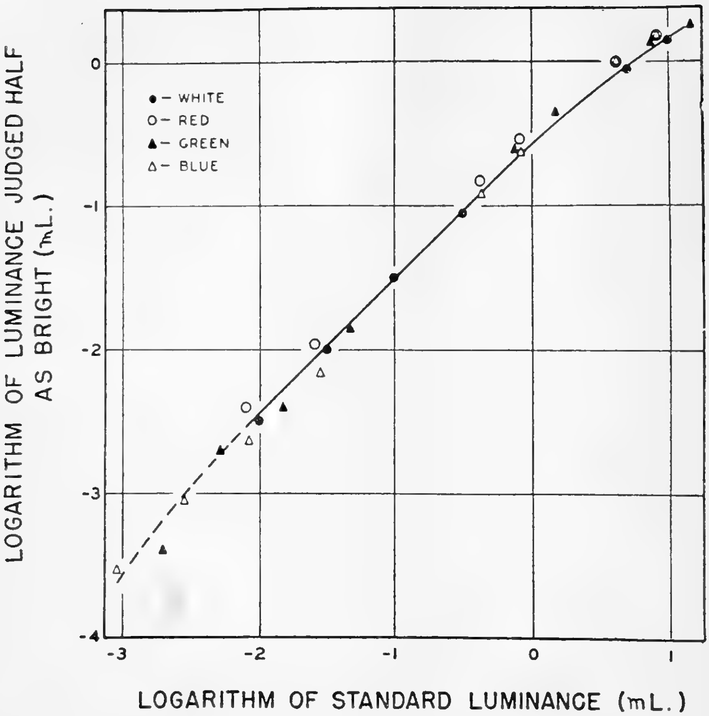

Fig. 46. Luminance of a test area judged to be half as bright as another test area. (After Hanes, **33**)

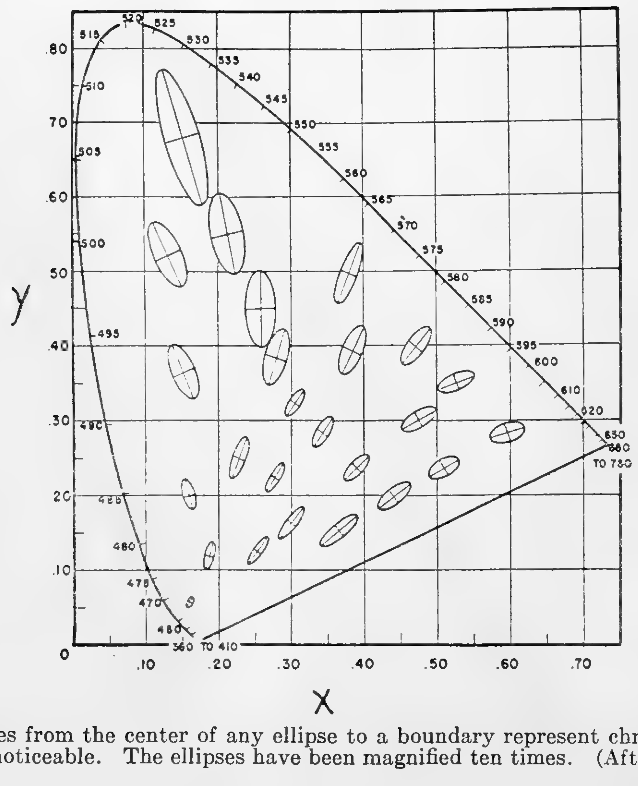

Fig. 47. Distances from the center of any ellipse to a boundary represent chromaticity differences which are equally noticeable. The ellipses have been magnified ten times. (After MacAdam, **63**)

The important point for purposes of this discussion is that changes in dominant wavelength and changes in colorimetric purity are not equally effective in arousing changes in color sensations. In some places on this chart the eye is more sensitive to dominant wavelength than to colorimetric purity (e.g., around 580 and 525 mμ). In other locations, it seems to be more sensitive to differences in colorimetric purity (e.g., around 485 mμ). This is one approach to the problem of measuring color sensations. Still others are possible.

Research Needed. This very brief discussion may be sufficient to show that the magnitudes of our sensations can be related systematically to the magnitudes used to measure light and color. In some ways, the former are the more important magnitudes because they are the ones which largely determine our reactions. Psychologists have just barely begun to explore this field and fruitful problems are here for the asking. Many important applications will undoubtedly await the accumulation of basic data in this area. As an illustration, it might be possible to use colors in a quantitative scale to represent the third dimension on visual displays. Before any such application as this can be made, however, it will be necessary to extend still further the basic work started by MacAdam.

References

1. Bitterman, M. E., & Soloway, E. The relation between frequency of blinking and effort expended in mental work. *J. exp. Psychol.*, 1946, **36**, 134–136.
2. Blackwell, H. R. Contrast thresholds of the human eye. *J. opt. Soc. Amer.*, 1946, **36**, 624–643.
3. Blanchard, J. The brightness sensibility of the retina. *Phys. Rev.*, 1918, **11**, 81–99.
4. Blondel, A., & Rey, J. Sur la perception des lumières brèves á la limite de leur portée. *J. Phys.*, Paris, 1911, **1**, 530–550.
5. Bouma, P. J. Illumination and black-outs. *Philips Tech. Rev.*, 1939, **4**, 15–19.
6. Chapanis, A. *Night vision tests and the night vision testing program in the Army Air Forces.* AAF Materiel Command, Aero Med. Lab. Memo. Rep. No. ENG-49-695-37C. 20 July 1944. 25 Pp.
7. Chapanis, A. The dark adaptation of the color anomalous. *Amer. J. Physiol.*, 1946, **146**, 689–701.
8. Chapanis, A. The dark adaptation of the color anomalous measured with lights of different hues. *J. gen. Physiol.*, 1947, **30**, 423–437.
9. Chapanis, A., Rouse, R. O., & Schachter, S. The effect of intersensory stimulation on dark adaptation and night vision. *J. exp. Psychol.* (In press.)
10. Cobb, P. W. The effect on foveal vision of bright surroundings—II. *Psychol. Rev.*, 1914, **21**, 23–32.
11. Cobb, P. W. Vision and the brightness of surroundings. *Trans. Ill. Engng. Soc.*, 1916, **11**, 372–388.
12. Cobb, P. W., & Geissler, L. R. The effect on foveal vision of bright surroundings. *Psychol. Rev.*, 1913, **20**, 425–447.
13. Cobb, P. W., & Moss, F. K. The effect of brightness on the precision of visually controlled operations. *J. Franklin Inst.*, 1925, **199**, 507–512.
14. Cobb, P. W., & Moss, F. K. Lighting and contrast. *Trans Illum. Engng. Soc.*, 1927, **22**, 195–204.
15. Cobb, F. W., & Moss, F. K. The four variables of the visual threshold. *J. Franklin Inst.*, 1928, **205**, 831–847.
16. Committee on Colorimetry. The concept of color. *J. opt. Soc. Amer.*, 1943, **33**, 544–554.
17. Committee on Colorimetry. Physical concepts: radiant energy and its measurement. *J. opt. Soc. Amer.*, 1944, **34**, 183–218.
18. Committee on Colorimetry. The psychophysics of color. *J. opt. Soc. Amer.*, 1944, **34**, 245–266.
19. Committee on Colorimetry. Quantitative data and methods for colorimetry. *J. opt. Soc. Amer.*, 1944, **34**, 633–688.
20. Connor, J. P., & Ganoung, R. E. An experimental determination of the visual thresholds at low values of illumination. *J. opt. Soc. Amer.*, 1935, **25**, 287–294.
21. Craik, K. J. W., & Vernon, M. D. Perception during dark adaptation. *Brit. J. Psychol.*, 1942, **32**, 206–230.
22. Crittenden, E. C., & Richtmyer, F. K. An "average eye" for heterochromatic photometry, and a comparison of a flicker and an equality-of-brightness photometer. *Trans. Illum. Engng. Soc.*, 1916, **11**, 331–356.
23. Crouch, C. L. The relation between illumination and vision. *Illum. Engng.*, 1945, **40**, 747–784.
24. Ferree, C. E., & Rand, G. The effect of variation of visual angle, intensity, and composition of light on important visual functions. *Trans. Illum. Engng. Soc.*, 1922, **17**, 69–102.
25. Ferree, C. E., & Rand, G. Visibility of objects as affected by color and composi-

tion of light. Part I. With lights of equal luminosity or brightness. *Person. J.*, 1931, **9,** 475–492.

26. Ferree, C. E., & Rand, G. Visibility of objects as affected by color and composition of light. Part II. With lights equalized in both brightness and saturation. *Person. J.*, 1931, **10,** 108–124.
27. Fisher, M. B. The relationship of the size of the surrounding field to visual acuity in the fovea. *J. exp. Psychol.*, 1938, **23,** 215–238.
28. Gibson, K. S., & Tyndall, E. P. T. Visibility of radiant energy. *Sci. Pap. Bur. Stand.*, 1923, **19,** No. 475.
29. Giese, W. J. The interrelationship of visual acuity at different distances. *J. appl. Psychol.*, 1946, **30,** 91–106.
30. Gordon, D. A. The relation between the thresholds of form, motion, and displacement in parafoveal and peripheral vision at a scotopic level of illumination. *Amer. J. Psychol.*, 1947, **60,** 202–225.
31. Haig, C. The course of rod adaptation as influenced by the intensity and duration of pre-adaptation to light. *J. gen. Physiol.*, 1941, **24,** 735–751.
32. Hammond, E. C., & Lee, R. H. *The reliability and validity of several night vision tests for the selection of Army truck drivers.* Div. Ind. Hyg., Nat. Inst. Health, unnumbered report. Feb. 1942. 67 Pp.
33. Hanes, R. M. A scale of subjective brightness. *J. exp. Psychol.* (In press)
34. Hanes, R. M., & Williams, S. B. Visibility on cathode-ray tube screens: the effects of light adaptation. *J. opt. Soc. Amer.*, 1948, **38,** 363–377.
35. Hecht, S. The instantaneous visual threshold after light adaptation. *Proc. nat. Acad. Sci.*, Wash., 1937, **23,** 227–233.
36. Hecht, S., Haig, C., & Chase, A. M. The influence of light adaptation on subsequent dark adaptation of the eye. *J. gen. Physiol.*, 1937, **20,** 831–850.
37. Hecht, S., & Hsia, Y. Dark adaptation following light adaptation to red and white lights. *J. opt. Soc. Amer.*, 1945, **35,** 261–267.
38. Hecht, S., Ross, S., & Mueller, C. G. The visibility of lines and squares at high brightness. *J. opt. Soc. Amer.*, 1947, **37,** 500–507.
39. Hecht, S., & Williams, R. E. The visibility of monochromatic radiation and the absorption spectrum of visual purple. *J. gen. Physiol.*, 1922, **5,** 1–33.
40. Hill, N. E. G. The recognition of coloured light signals which are near the limit of visibility. *Proc. phys. Soc.*, Lond., 1947, **59,** 560–574.
41. Johnson, H. M. Speed, accuracy and constancy of response to visual stimuli as related to the distribution of brightnesses over the visual field. *J. exp. Psychol.*, 1924, **7,** 1–44.
42. Jones, L. A. The historical background and evolution of the colorimetry report. *J. opt. Soc. Amer.*, 1943, **33,** 534–543.
43. Keil, F. C., Jr. *The comparison of the Hecht Adaptometer and the Miles Illuminous Disc Adaptometer.* AAF Sch. Aviat. Med., Rep. No. 1, Project No. 17. 31 Mar. 1942.
44. Kekceev, K. C. On the action of nonadequate stimuli on receptors. *C. R. Acad. Sci. U. R. S. S.*, 1937, **14,** 495–497.
45. Kekcheev, K. K. Methods of accelerating dark adaptation and improving night vision. *War Med.*, 1945, **8,** 209–220.
46. Kekcheyev, K. Expediting visual adaptation to darkness. *Nature*, 1943, **151,** 617–618.
47. Kektchéeff, K. C., & Ostrovsky, E. P. Sur la définition des vibrations aériennes de fréquence ultra-sonore par la mesure des seuils optiques. *C. R. Acad. Sci. U. R. S. S.*, 1941, **31,** 370–372.
48. Knoll, H. A., Tousey, R., & Hulburt, E. O. Visual thresholds of steady point sources of light in fields of brightness from dark to daylight. *J. opt. Soc. Amer.*, 1946, **36,** 480–482.
49. Kohlrausch, A. Untersuchungen mit farbigen Schwellenprüflichtern über den Dunkeladaptationsverlauf des normalen Auges. *Arch. ges. Physiol.*, 1922, **196,** 113–117.
50. Lasareff, P. P. Théorie thermique des changements de la sensibilité visuelle périphérique produits par les causes géophysiques. *C. R. Acad. Sci. U. R. S. S.*, 1937, **14,** 275–278.
51. Lash, J. D., & Prideaux, G. F. Visibility of signal lights. *Illum. Engng.*, 1943, **38,** 481–492.
52. Livingstone, P. C. Visual problems of aerial warfare. *Lancet*, 1944, **247,** 33–39.
53. Low, F. N. The development of peripheral visual acuity during the process of dark adaptation. *Amer. J. Physiol.*, 1946, **146,** 622–629.
54. Low, F. N. The peripheral motion acuity of 50 subjects. *Amer. J. Physiol.*, 1947, **148,** 124–133.

55. Lowry, E. M. The effect of hue on dark adaptation. *J. opt. Soc. Amer.*, 1943, **33**, 619–620.
56. Luckiesh, M. The influence of spectral character of light on the effectiveness of illumination. *Trans. Illum. Engng. Soc.*, 1912, **7**, 135–152.
57. Luckiesh, M. *Light, vision and seeing.* New York: D. Van Nostrand, 1944.
58. Luckiesh, M., & Moss, F. K. Light, vision and seeing. In: O. Glasser (Ed.) *Medical Physics.* Chicago: Year Book Publishers, 1944. Pp. 672–684.
59. Luckiesh, M., & Taylor, A. H. A summary of researches in seeing at low brightness-levels. *Illum. Engng.*, 1943, **38**, 189–207.
60. Ludvigh, E. Visibility of the deer fly in flight. *Science*, 1947, **105**, 176–177.
61. Ludvigh, E. The visibility of moving objects. *Science*, 1948, **108**, 63–64.
62. Lythgoe, R. J. The measurement of visual acuity. *Spec. Rep. Ser., med. Res. Coun.*, Lond., 1932, No. 173. 85 pp.
63. MacAdam, D. L. Visual sensitivities to color differences in daylight. *J. opt. Soc. Amer.*, 1942, **32**, 247–274.
64. Mandelbaum, J., & Rowland, L. S. *Central and paracentral acuity at different levels of illumination.* AAF Sch. Aviat. Med. Res. Rep. No. 1, Project No. 220, 23 June 1944.
65. Matthews, B. H. C., & Luczak, A. K. *Some factors influencing dark adaptation.* Flying Personnel Research Committee Report No. 577, RAF Physiological Laboratory, Farnborough, June, 1944. 33 pp.
66. Miles, W. R. Night vision: flying demands light sensitivity and form acuity. *Yale sci. Mag.*, 1943, **18**, 10–11, 28, 30.
67. Moon, P. A system of photometric concepts. *J. opt. Soc. Amer.*, 1942, **32**, 348–362.
68. Moon, P. *The scientific basis of illuminating engineering.* New York: McGraw-Hill, 1936.
69. Moon, P., & Spencer, D. E. Visual data applied to lighting design. *J. opt. Soc. Amer.*, 1944, **34**, 605–617.
70. Müller, H. K. Über den Einfluss verschieden langer Vorbelichtung auf die Dunkeladaptation und auf die Fehlergrösse der Schwellenreizbestimmung während der Dunkelanpassung. *Arch. Ophthal.*, Berlin, 1931, **125**, 624–642.
71. Olenski, Z., & Goodden, N. W. *Clearness of view from day-fighter aircraft.* Farnborough Royal Aircraft Establishment, Report No. Aero 1862. Oct. 1943.
72. Østerberg, G. Topography of the layer of rods and cones in the human retina. *Acta ophthal., Kbh.*, 1935, **13**, Suppl. VI, Pp. 103.
73. Peckham, R. H. *The course of readaptation after exposure to white light.* Res. Rep., U. S. Naval Air Training Cntr., Naval Sch. Aviat. Med., Pensacola, Fla. 23 Feb. 1944. 14 pp.
74. Pinson, E. A., & Chapanis, A. *Tests of dark adaptation and night vision: their reliabilities and intercorrelations.* AAF Materiel Command, Aero Med. Lab. Memo. Rep. No. ENG-695-12-Q. 17 Dec. 1943. 19 Pp.
75. Reinhardt, H. Brightness units. *Illum. Engng.*, 1944, **39**, 521–533.
76. Rose, H. W., & Schmidt, I. Factors affecting dark adaptation. *J. Aviat. Med.*, 1947, **18**, 218–230.
77. Rowland, W. M., & Mandelbaum, J. *A comparison of night vision testers.* AAF Sch. Aviat. Med. Res. Rep. No. 1, Project No. 213, 26 Jan. 1944. 13 Pp.
78. Rowland, W. M., & Rowland, L. S. *Aspects of night visual efficiency.* AAF Sch. Aviat. Med., Rep. No. 1, Project No. 106. 1 May 1943.
79. Rowland, W. M., & Sloan, L. L. The relative merits of red and white light of low intensity for adapting the eyes to darkness. *J. opt. Soc. Amer.*, 1944, **34**, 601–604.
80. Semeonoff, B. Sensitivity of the dark-adapted eye during a prolonged period of observation. *Nature*, 1941, **147**, 454–455.
81. Serrat, W. D., & Karwoski, T. An investigation of the effect of auditory stimulation on visual sensitivity. *J. exp. Psychol.*, 1936, **19**, 604–611.
82. Shlaer, S., Smith, E. L., & Chase, A. M. Visual acuity and illumination in different spectral regions. *J. gen. Physiol.*, 1942, **25**, 553–569.
83. Sloan, L. L. The effect of intensity of light, state of adaptation of the eye, and size of photometric field on the visibility curve. *Psychol. Monogr.*, 1928, **38**, No. 1.
84. Sloan, L. L. Rate of dark adaptation and regional threshold gradient of the dark-adapted eye: physiologic and clinical studies. *Amer. J. Ophthal.*, 1947, **30**, 705–720.
85. Sulzman, J. H., Cook, E. B., & Bartlett, N. R. The reliability of visual acuity scores yielded by three commercial devices. *J. appl. Psychol.*, 1947, **31**, 236–240.

86. Taylor, A. H. Vision at low brightness-levels. *Illum. Engng.*, 1943, **38**, 89–98.

87. Thorne, F. C. The psychophysical measurement of the temporal course of visual sensitivity. *Arch. Psychol.*, N. Y., **1934**, No. 170. 66 pp.

88. Tinker, M. A. Validity of frequency of blinking as a criterion of readability. *J. exp. Psychol.*, 1946, **36**, 453–460.

89. Troland, L. T. Report of Committee on Colorimetry for 1920–21. *J. opt. Soc. Amer.*, 1922, **6**, 527–596.

90. Turlygin, S. J. Effect of electromagnetic centimetre waves on the central nervous system. *C. R. Acad. Sci. U. R. S. S.*, 1937, **17**, 19–21.

91. Wald, G. Human vision and the spectrum. *Science*, 1945, **101**, 653–658.

92. Wald, G., & Clark, A. B. Visual adaptation and chemistry of the rods. *J. gen. Physiol.*, 1937, **21**, 93–105.

93. Walls, G. L. Factors in human visual resolution. *J. opt. Soc. Amer.*, 1943, **33**, 487–505.

94. Walsh, J. W. T. *Photometry*. New York: D. Van Nostrand, 1926.

95. Walters, H. V., & Wright, W. D. The spectral sensitivity of the fovea and extrafovea in the Purkinje range. *Proc. roy. Soc.*, 1943, **131B**, 340–361.

96. Wentworth, H. A. A quantitative study of achromatic and chromatic sensitivity from center to periphery of the visual field. *Psychol. Monogr.*, 1930, **40**, No. 3.

97. Wertheim, T. Über die indirekte Sehschärfe. *Z. Psychol.*, 1894, **7**, 172–187.

98. Winsor, C. P., & Clark, A.-B. Dark adaptation after varying degrees of light adaptation. *Proc. nat. Acad. Sci.*, Wash., 1936, **22**, 400–404.

99. *IES lighting handbook: the standard lighting guide.* Published by The Illuminating Engineering Society, 51 Madison Avenue, New York, 1947.

100. *Illuminating engineering nomenclature and photometric standards. American Standard 1942.* Prepared under the direction of and issued by the Illuminating Engineering Society, 51 Madison Ave., New York. Pp. 42.

101. *Influence of color contrast on visual acuity.* Natl. Def. Res. Com., Off. Sci. Res. Dev. Rep. No. 4541. Prepared by the Eastman Kodak Co., 1944.

102. *Interconversion tables for photometric units.* Admiralty Res. Lab., Teddington, Middlesex, Rep. No. A. R. L./N. 1/0.288. Oct. 10, 1944.

103. *Studies in visual acuity.* Dep. of the Army, PRS Rep. No. 742. Prepared by the Staff, Personnel Research Section, The Adjutant General's Office, Washington: Govt. Printing Office, 1948.

Chapter 2

PRINTED MATERIALS, MAPS, AND CHARTS

MASON N. CROOK

Tufts College

Introduction

The use of maps, charts, and printed materials enlists a variety of sensory, perceptual, and motor functions on the part of the user. We are here concerned with the problem how to arrange conditions most advantageously for the performance of these functions. Solutions must often be a compromise among capacities of the organism, characteristics of the work situation, and technological limitations in the production of materia ls. Certain aspects of performance, such a s the increased difficulty of reading maps under colored light, can be predicted from a general knowledge of the visual receptor mechanism as set forth in Chapter 1. Most visual tasks involving the interpretation of symbols, however, are not capable of full theoretical analysis and must be approached empirically. Ordinary reading typifies this state of affairs and also illustrates many of the problems, particularly of methodology and interpretation, encountered in the investigation of other visual tasks. Therefore, a brief review of some representative samples of the extensive work on this topic will provide us with a helpful frame of reference. Much of the material can be treated in relation to three variables, typography, illumination, and fatigue.

Reading Printed Text

We can distinguish two general kinds of experimental situations: (1) those in which subjects are required to approximate normal reading of good quality book type, in relation to which certain functions are measured, and (2) those in which some form of threshold determination is made. Among the latter are determinations of the distance at which type can be read, of the effects of impairment of focus of a projected image, of discrimination in brief exposures, of reading limits with the Luckiesh-Moss "visibility" meter, and of acuity by classical methods. The significance of such measures for normal reading is at best inferential, and different methods have in some cases given clearly different results or have been open to different interpretations. (See, for example, Chapter 1 on visual acuity measures.) We shall, therefore, confine ourselves in this section to the more direct investigations of normal reading. Tests of legibility at a distance, which are relevant to the reading of wall placards, will be separately discussed.

Even within the area of normal reading studies, a conflict exists over ways of measuring results. Measures of reading speed by means of standardized tests are held by Tinker to be a satisfactory criterion. He is opposed by Luckiesh and Moss, who claim (1) that reading speed is a relatively insensitive measure, and (2) that it is a measure of output only, and takes no account of the cost to the organism of doing the work of reading. Proceeding on the assumption that some involuntary behavior concomitant with reading would be likely to indicate the effort expended, they have investigated a number of such concomitants, among them muscle tension, blink rate, and heart rate. Of these they have settled on blink rate as the most usable (**30,** pp. 115–121).[1] In a

[1] When research programs have been summarized in book form, the book references only will be cited. References to selected experimental articles not covered in such summaries, and to the more relevant surveys and reviews, will be given. Much of the material on typographical

study by Hoffman (**26**) of reading continued for four hours, blink rate, together with other aspects of eye movements, showed significant changes. But several investigators, evaluating the blink rate experimentally, have failed to find it a satisfactory measure (**4, 8, 9, 31, 52, 53**.) It is thus seen that, though blink rate may in some circumstances reflect the conditions of reading, the lack of consistent results when this index is used by different workers makes it difficult to evaluate any single set of data. This controversy has been covered in an analytic review by Bitterman (**6**).

The well-known capacity of the human organism to compensate for unfavorable conditions, and especially the tendency to do so in brief test situations that are recognized as tests by the subjects, often results in uniformly high performance scores when conditions are widely varied. This gives rise to two problems. The first is the problem of getting an index of the energy cost per unit of work output. In the case of visual work the measurement of associated involuntary responses is a step, at present of uncertain value, in this direction. The need for dependable measures for this purpose is generally recognized.

The second problem is that of determining the consequences to the organism of a given level of energy cost. The implicit assumption often made is that the optimum cost for a given work output is the minimum cost. This assumption seems safe enough when the task is arduous or when the subject is working near the limit of his tolerance or endurance. But when the work is light and other conditions are not conspicuously unfavorable, the situation is less clear. There are three possibilities: (1) that the more costly way of doing a task produces cumulative impairment over a period of a day or several days, (2) that the additional energy loss and discomfort is negligible, relative to the total demands on the organism and to the capacities of the particular organs involved, or (3) that the higher energy output is favorable in terms of some secondary criterion such as physical tonus or health. When immediate cost can be satisfactorily measured, the consequences to the organism will need to be studied as a separate problem.

Typographical Factors

With these considerations in mind we turn to studies of typographical variables. Such studies have quite generally been subject to the limitation that the reading periods involved were brief. More studies of prolonged reading are needed to verify the assumption now tentatively made that any effects appearing in brief reading periods will appear to at least as great an extent in longer periods.[2]

A few words about details of typography. The unit of measure for type sizes is the "point." A point is 0.0138 inch, or 0.351 millimeter. Three-point type is about the smallest that can be read at 14 inches under favorable conditions. With occasional exceptions, newspapers are printed in 7- or 8-point, jounals in 9- to 11-point, and textbooks in 10- to 12-point. The point size gives exact dimensions for the type body only. Dimensions of the letter mounted on the body will vary with the type face. Line width (the printer's term for line length) is measured in picas. One pica is 12 points (0.166 in. or 4.21 mm). Single column textbook lines are usually between 19 and 24 picas. Leading, the space between lines, varies somewhat with type size and other factors, but most textbooks are set with 2-point leading. The present volume is prepared in 10 point type, with 16 picas to the column and 2 point leading.

Many type faces have been designed, but a relatively small number of them are widely favored for ordinary printing in books and journals. Type faces differ, not only in

---

factors has been covered in a review for the Air Force by Young (**59**).

[2] Reading periods up to 10 or 15 minutes are here classified as "short," and periods from one-half hour to several hours as "long."

letter height, but in height-width ratio and in thickness of stroke (boldness). In a given type face, a smaller point size may not be a photographic reduction of a larger, since certain letters may differ in details of design, and the reductions in letter height and in thickness of stroke may not be in proportion to each other nor to the reduction in point size. Type designers have resorted to these dimensional disparities to preserve certain perceptual uniformities as pleasing appearance, legibility, and style.

A comprehensive investigation of typographical factors carried out by Paterson, Tinker, and associates has been brought together in their book, *How to Make Type Readable* (**33**). Considerable work by Luckiesh and Moss is reported in *The Science of Seeing* (**29**) and *Reading as a Visual Task* (**30**).

Paterson and Tinker gave standard speed tests, taking $1\frac{3}{4}$ minutes, under classroom conditions, to school and college students, in groups of 80 or 90 to several hundred, totaling 33,000. Luckiesh and Moss reported results usually in terms of blink rate, sometimes in terms of reading speed when subjects were reading at their normal rate, and occasionally in terms of reading speed when subjects were reading at their maximum rate. Experimental groups contained typically 30 or 40 subjects, lighting was controlled, and sometimes retests were made. The test period was usually five minutes.

Kinds of Type

With other variables held constant, style of type face appears to be a minor factor in ordinary reading. Paterson and Tinker (**33,** p. 16) found a difference of less than three percent in reading speed among eight type faces in common use, but American Typewriter reduced speed about five percent and Old English about 14 percent. Luckiesh and Moss (**30,** p. 163) in testing three type faces found insignificant differences in reading speed, but up to 14 percent differences in blink rate. The relative blink rates, however, were in the order of relative boldness for the three type faces.

Boldness

In another study, Luckiesh and Moss (**30,** p. 171) found some differentiation beween degrees of boldness, in a particular type face, in terms of reading speed and blink rate, the optimum boldness being somewhat heavier than that of the average type face in common use. Paterson and Tinker (**33,** p. 27) found boldness not to be a significant variable.

Italics and Capitals

In comparison with ordinary lower case, Paterson and Tinker (**33,** pp. 21–23) found that italicized material was read about three percent more slowly and material set in all capitals about 12 percent more slowly.

Type Size, Line Width, and Leading

On the question of type size, the several investigators emphasize the interaction with line width and leading. When one factor alone is varied experimentally, the results tend to be specific for the conditions of the experiment. Therefore, more helpful conclusions can be drawn from experiments which include combinations of variables. Type sizes from 9- to 12-point were found by Paterson and Tinker (**33,** p. 81) to be read with about equal speed when each was set in its optimum line width at 2-point leading. In a somewhat analogous experiment with the blink-rate criterion, Luckiesh and Moss (**30,** p. 150) found generally similar results. A further experiment by Paterson and Tinker (**33,** p. 80) enabled them to define, for each type size, certain safety zones of line width and leading within which reading speed was not impaired. For 6- or 8-point type set solid (0 leading), line width could be varied from 14 to something under 28 picas. Increasing leading through two points tended to extend the line width range. Increasing type size tended to extend the range and raise the lower limit slightly, so that for 12-point type leaded two points, the safety zone was from 17 to something under 41 picas.

Color and Brightness Relations

Paterson and Tinker (**33,** p. 113) and Taylor (**43**) found that white printing on black paper was read eight to 10 percent less rapidly than black on white. Variations in color of ink and paper were found by both Paterson and Tinker (**33,** Chap. 10), using reading speed, and Luckiesh and Moss (**30,** p. 224), using reading speed and blink rate, to produce large variation in legibility. Those combinations which showed the greater brightness contrasts were in general the more legible.

Paper Surface

Paterson and Tinker (**33,** pp. 132 ff.) reported that materials printed on glossy and dull paper were read with equal speed. This finding was to some extent confirmed by Stanton and Brutt (**41**), using the same general procedure. For brief reading periods, the possible glare effect of glossy paper may be outweighed by the sharper printing impression obtained.

Arrangement on the Page

To the average reader, the printing on a textbook page with average margins appears to cover about 75 percent of the page area. This effect is illusory, the actual area covered being only about 50 percent. The unused space seems to serve mainly an aesthetic function, since Paterson and Tinker (**33,** p. 98) have found that pages with $\frac{7}{8}$-inch and with $\frac{1}{16}$-inch margins are read equally rapidly.

Methods of separating double columns vary from the use of an inter-column rule with very little space to the use of a two-pica space with no rule. Paterson and Tinker (**33,** p. 105) found that inter-column arrangements, even including the rather extreme case of a $\frac{1}{2}$-pica space with no rule, had no effect on reading speed. But the breaking of print into short paragraphs was found to facilitate reading (**33,** p. 106 ff.)

Further Combinations of Factors

In a final study in their series, Paterson and Tinker (**33,** Chap. 12) found that when several unfavorable typographical conditions were presented together, their separate impairments did not combine in a simple additive way. This result further emphasizes the complexity of the interactions among the many typographical variables.

In general, these findings on typographical factors as evaluated in brief test periods under reasonably good reading conditions show little serious disagreement, in spite of conflicts of point of view among the major investigators. Those factors which make little difference are type face, at least within the range of those commonly used for printing, italics, paper stock, margins, and inter-column spacing. Those which have a substantial effect are the use of special type faces as American Typewriter or Old English or of all capitals, paragraphing, white printing on a black ground, and colors which make for low brightness contrast. The evidence on boldness is conflicting. Interactions among the factors are important.

Illumination

Certain principles of illumination, such as the desirability of good diffusion and the elimination of glare spots and areas of high brightness contrast, are generally accepted and need no elaboration. But on the question of desirable amounts of illumination there are sharp differences of opinion, here again between Luckiesh and Moss on one hand and Tinker on the other. The former recommend increasing amounts of light for progressively more difficult work up to a minimum of 100 foot-candles for very exacting tasks (**30,** p. 330). Illuminating engineers have tended to follow Luckiesh and Moss. Tinker (**54**) has criticized these standards, particularly in the upper part of the range, on the following grounds, among others:

1. The blink rate and allied criteria are questionable.

2. Luckiesh and Moss present little or no data in any terms for the region between 10 and 100 foot-candles.

3. The recommendations rest in part on estimates from various threshold measures which have not been validated for actual visual work activities.

In contrast to the Luckiesh and Moss minimum of 100 foot-candles for the most severe visual tasks, Tinker considers a maximum of 50 to be adequate. For a critique of this controversy, see Bitterman's review mentioned above (**6**).

Brief Reading Periods

Reading is usually treated as a visual task of intermediate difficulty. Returning to the situation of brief reading periods, we find that Luckiesh and Moss (**30,** p. 330) recommend 10 to 20 foot-candles, while Tinker (**54**), not differentiating between long and short periods, recommends 10 to 15 foot-candles. Tinker's recommendations are intended to include a safety margin. His own experiments (**50**), covering a range from 0.1 to 53.3 foot-candles, showed no improvement above three foot-candles when the subjects were in every case well adapted to the reading illumination level. Rose and Rostas (**39**) found no significant improvement in speed or comprehension from two to 55 foot-candles. Luckiesh and Moss (**30,** pp. 107, 114, 190) report several sets of data, which show, among other things, that reading speed increases above 10 foot-candles, but only at the rate of five percent for the interval between 10 and 100. Tinker's placing of the critical level (the point above which further gains are inconsequential) at three foot-candles when the subjects are adequately adapted seems to fit the central tendency of these several sets of data reasonably well.

Sustained Reading

For periods of sustained reading, Luckiesh and Moss recommend 20 to 50 foot-candles (**30,** p. 330) as against Tinker's 10 to 15 (**54**). Luckiesh and Moss (**30,** p. 114) report that, for periods of 30 minutes to an hour, reading speeds increase as illumination increases from one through 10 to 100 foot-candles, the maximum difference being about seven percent.

A number of experiments have been set up in such a way that illumination could be related to time trends in some aspect of the reading process or its concomitants. Tinker (**50**) found that clearness of vision was impaired by two hours of reading at illuminations below three foot-candles, but was not affected at that level or above. Luckiesh and Moss (**30,** p. 360) found that the power of the convergence response of the eyes was impaired 23 percent by one hour of reading under one foot-candle, but only seven percent under 100 foot-candles.

Certain sets of data are in open conflict. Luckiesh and Moss found, for periods of sustained reading, that blink rate increased and heart rate decreased less rapidly under the higher illuminations (**30,** pp. 107, 353). Bitterman (**4**) and McFarland, Holway, and Hurvich (**31**), on the other hand, found inversions in the blink rate–illumination relationship, with large individual differences in trends, and McFarland, Knehr, and Berens (**32**) found no relation between heart rate and illumination. These several authors interpret their results, not as disproving illumination effects, but as invalidating the blink-rate criteria. Bitterman (**4**) found heart rate changes generally in accord with those of Luckiesh and Moss, but on the grounds of other evidence rejected the interpretation that this indicated greater visual effort at the lower illuminations.

In looking for points on which there is no conspicuous disagreement among these diversified data, we find, first an indication of definite impairment below three foot-candles, and second, the fact that, as illumination increases, performance scores improve at a decreasing rate. On the question of a critical level the evidence is conflicting. If we minimize the controversial data on blink rate and heart rate, the remaining findings, though not in good agreement with each other, suggest a value somewhat above the three foot-candles for short reading periods—perhaps 10 foot-candles. If we include a common-sense safety margin of 10 foot-candles, practical recommendations based only on those studies in which the subjects actually did normal reading under good conditions would come to 13 foot-candles for brief, and 20 foot-candles for long, reading periods.

An extreme illumination situation is met in the use of faint red light to preserve dark adaptation. Reading of books and maps is alleged not to be difficult under this condition (Keil, **28**) but quantitative information is meager.

Fatigue

Subjective impressions, and analogy with other forms of work, support the hypothesis that sustained reading is accompanied by significant amounts of fatigue. However, the phenomenon has proved peculiarly elusive under experimental attack. The monograph by McFarland, Holway, and Hurvich (**31**) reported an extensive exploratory study of the mechanisms of visual fatigue which failed to find grounds for a consistent general theory. Several other studies mentioned above in the discussion of problems of illumination found that reading was accompanied by changes in certain functions (**4; 30,** pp. 107, 360; **31, 50**), but usually under only the lower illuminations or in terms of the measures most subject to dispute. The most systematic positive findings are those reported by Hoffman (**26**), who took electrical records of the eye movements of 30 subjects as they read continuously for four hours under 10.5 foot-candles. From an analysis of the records, indices were computed for blinks, fixations, regressions, lines read (speed), fixations per line, regressions per line, and variability of fixations. By the end of the four-hour reading period, all of the indices showed significant changes, except regressions, which fluctuated irregularly. Carmichael and Dearborn (**10**),[3] on the other hand, in a very similar experiment in which more extensive data were gathered, obtained negative results. High school and college students read book print for six-hour periods under 16 foot-candles. The entire procedure was repeated with microfilm, results for which are discussed below. For the book reading, neither eye movement records analyzed for the same indices as in the Hoffman experiment, nor comprehension measures obtained on the average every 25 pages, showed reliable changes over the six hours. Acuity, measured before and after, showed no changes that could be clearly assigned to the effect of reading. Heart and brain-wave records, not fully analyzed, showed no obvious relation to the reading task. Subjective impressions gathered at the end of reading included a number of reports of discomfort, but these the authors interpreted as related rather to the restrictions in posture than to the reading itself. It appears that any physiological changes produced by the large amount of ocular work involved in prolonged reading can readily be compensated for by the subjects, and the degree of discomfort does not seem to be excessive.

The foregoing comments on typography, illumination, and fatigue in ordinary reading, taken as a whole, permit the conclusion that optimal conditions are within the range of current facilities and practices. New departures are not indicated, with the possible exception of higher illumination, the need for which is under debate. From the standpoint of methodology, certain considerations relevant to the evaluation of other visual situations can be formulated.

First, since variables tend to interact, com-

[3] This reference contains a bibliography of 409 titles.

prehensive studies are to be preferred over studies of variables in isolation.

Second, the energy cost of reading is an obscure factor, the consequences of which have not been demonstrated by the evidence to date. The matter can hardly be left with the implication that therefore the cost must be negligible, because the problem may be accentuated for visual work under difficult conditions. More dependable measures in this area are needed.

Third, the emphasis in these studies has been on optimal conditions. Unfavorable reading situations likely to be encountered in practice have received much less systematic attention.

Inferior Materials

Examples of unfavorable reading conditions are the low quality paper used in pulp magazines and paper bound books, and the low intensity of illumination needed at some posts to protect the reader's dark adaptation. Scattered studies of various materials of inferior quality, such as newspaper print, typed carbons and originals, stencil duplicated copy, and comic book lettering, have shown a need for higher illumination levels (Tinker, **51**). With such materials there is also an impairment of reading speed and large variability in "visibility" meter measures (Leckiesh and Moss, **30**, Chap. IX). Thus, results of fatigue studies which employ good quality print cannot safely be extrapolated to such materials.

Microfilm

Microfilm presents a marginal reading situation which might at some future time become the norm. McFarland, Holway, and Hurvich (**31**, pp. 76 ff.) found no significant difference in blink rate between microfilm and book reading, but their subjects preferred to read the book. When given a choice of two illumination levels on the microfilm surface, the subjects chose the lower (**31**, p. 189). In the Carmichael and Dearborn study (**10**) eye movement measures differentiated between microfilm and normal book reading for high school but not for college students. Subjective reports of boredom and fatigue were in general more frequent for microfilm. This the authors attributed to the more awkward sitting position for microfilm reading. It appears from these results that any departure from the normal reading of good quality book type introduces new factors requiring separate evaluation.

Perception of Print at a Distance

The reading of placards and posters differs from book reading in important respects. There is relatively little content to aid perception. Distance often puts the material near the threshold of discrimination, and because the material is usually brief, the "work" component is reduced to a minimum. This situation is approximated in the experimental procedure of testing legibility by determining the distances at which words and letters can be identified. Results by this method agree with those in terms of reading speed in the following respects:

1. Relative legibility with different color combinations is roughly proportional to their brightness contrasts (Preston, Schwankl, and Tinker, **35**; Sumner, **42**).
2. Type of paper surface makes no difference (Roethlein, **38**; Webster and Tinker, **56**).

The two methods yield quite different results when type form or style is the variable:

1. Though lower case is read more rapidly, capitals can be identified at greater distances (Tinker, **49**), possibly because capitals cover larger areas.
2. Different type faces have quite different thresholds. Roethlein (**38**) presented letters singly and in combinations in a large number of type faces. She concluded that details of design are the least important, and relative boldness of the type faces one of the most important, factors, and that combinations of adjacent letters of similar outline, such as d and b, impair discrimination. Webster and

Tinker (**57**), using the same type faces previously tested for reading speed, found substantial differences in distance thresholds. American Typewriter, especially poor for reading speed, was especially good by the distance criterion. They attributed the distance results to such factors as relative size and boldness, simplicity versus complexity of design, area of white spaces, and relative dependence on hair lines.

Such comparisons suggest that the reading of continuous text is a highly integrative process to which context, word form, and detail of letter design all contribute; that the recognition of words at a distance is less integrative in that it makes less use of context; and that the recognition of letters at a distance is the most analytic of the three, being dependent entirely on detail of design.

Numerals

The reading of printed numerals under ordinary working conditions stands at the analytic end of the continuum set up in the preceding section. Context and word-form cues are lacking, and correct identification of each digit is usually critical. Integrative habits which enhance efficiency in reading verbal text may be disadvantageous with numerals. For example, it has been found that familiar number combinations such as 1492 tend to be read like word units (Rebert, **37**). This tendency could cause a combination similar to a familiar one to be misread.

It could be expected that the more analytic the reader's function, the better the chance of predicting relative legibility from knowledge of the visual mechanisms, and also that the nearer to a threshold in any respect, the better the chance of such predictions. Dial numerals under night illumination constitute the extreme example. (See Chapter 3 below.) Among the pointers and leads in this connection, we have Craik's (**15**) data on the relation between numeral size and illumination for constant legibility under night conditions, Berger's (**3**) experimentally verified prediction from acuity data, that the optimum stroke width would be greater for black digits on a white ground than for the reverse, and Sanford's (**40**) earlier analysis of the design details that differentiate the lower case letters. Such devices as the Luckiesh-Moss "visibility" meter, and the Weiss focal variator (**58**), which gives thresholds in terms of impairment of focus, might prove more helpful for numeral legibility than for verbal text. A survey of the sensory factors in digit design has been made for the Special Devices Center by Gleason (**23**). He concluded that such factors as size of critical detail, brightness, and contrast are well enough understood, but that more work is needed on stroke width and configuration.

The helpfulness of general theory does not emancipate us from empirical checks, especially where complex interactions may play a part. No integrated program on numeral legibility comparable to that of Paterson and Tinker on typography in ordinary reading has been carried out, but scattered studies on various aspects of the problem can be cited.

Typographical Factors in Digit Legibility

Type Size

Type size is one of the factors being investigated by Crook, Hoffman, Kennedy, and associates in a series of studies on numeral legibility currently in progress at Tufts College (**16, 17, 18, 19**). In some experiments the reading task has involved a sequence of "same-different" judgments on pairs of digits. In others it involved a sequence of mental addition problems. One- to three-minute trials were separated by comparable rest intervals. The material was printed in Modern type face. With the numerals in the normal working position, type sizes from 6- to 11-point were not differentiated in time or error scores. With the material inverted, however, significant differences appeared in both measures, small type being read less well (**17**). The disadvantage of small type for reading the digits inverted suggests, in the terms proposed above, that this operation is even more analytic than that of reading them upright.

Type Face

Tinker (**48**) found that modern and old style digits were read with equal speed and accuracy under normal reading conditions. When presented in groups at a distance, the old style digits were easier to perceive. As the two sets are not very different in details of form, the author attributed the advantage of the old style to the fact that some of the digits in this set project above or below the line. However, adjacent rising digits tended to confuse each other, as Roethlein (**38**) found for letters.

White on Black

It will be recalled that, for verbal material, white on black is less legible than black on white. In the Tufts study with printed numerals (**16**), white on black was read with the same speed and accuracy as black on white. Dunlap (**22**), in an investigation of license plates, concluded that dark numbers on a light ground were better, but this result, according to Berger (**3**) was probably a function of stroke widths used in the study. Berger's own data show that white license-plate numbers on a dark ground are more perceptible than black on white when each set is prepared in its optimal stroke width.

Digit Configuration

The present designs of the digits, like those of letters, were determined to a large extent by convenience in making the strokes before printing was invented. With printing, type makers have modified numeral details, perhaps more often for aesthetic reasons than for ease of discrimination. As a result, in some type faces the three and five are differentiated by one small vertical segment only. Sixes and nines are sometimes made like zeros with small tails, and are difficult to identify under near threshold conditions. Work by Tinker (**44**) and by Berger (**3**) shows that the legibility of the digits varies considerably. If our conventional number forms could be disregarded, it is probable that a much improved set of symbols could be devised. Short of that, we have Berger's (**3**) work in improving the configuration of license-plate numerals, in which he found the angles of the slanting lines to be one important factor. Work on this general problem with dial-type digits is under way at Tufts (**19**).

Spacing

The legibility of digits presented in groups is different in both relative and absolute terms from that of digits presented singly (Tinker, **48**; Berger, **3**). Berger found that spacing could be used to some extent to control the thresholds. It is possible, therefore, that printed numerals which usually appear in groups may require different designs from dial numerals which often appear singly or in limited combinations.

For further discussion of the influence of some of these variables on numeral legibility, the reader is referred to the section in Chapter 3 which deals with instrument marking.

Other Factors in Digit Legibility

Illumination

Atkins (**2**) found that time scores on a number cross-out test differed by less than four percent for five illuminations ranging from 9.6 to 118.0 foot-candles. In the Tufts studies (**17**), time and error scores in reading printed numerals were not significantly different under brightnesses from one to 15 foot-lamberts. (Again, see Chapter 3 for related data on instrument dial numerals.)

Vibration

In the Tufts experiments (**16, 17, 18, 19**) vibration was introduced into the visual field by optical means. With secondary conditions moderately severe, a double amplitude of vibration of 0.02 inch was sometimes found to produce a measureable impairment in time or error scores. The effect varied with such characteristics of the vibration as frequency and pattern.

Interactions

As indicated by some of the results above, it was found in the Tufts experiments that type size, illumination, and vibration amplitude could be separately varied over considerable ranges without seriously affecting performance. But when the factors were varied together over the same ranges, unfavorable values of the three in combination produced marked impairment (**17, 19**).

Fatigue

The reading of numerals imposes demands on attention, and probably on eye movements, that are different from those made by the reading of verbal text. It might be of some importance to know whether numeral reading is more fatiguing. Available evidence is indirect. In the Tufts experiments the schedule was designed to minimize fatigue, but some indication of impairment during a working period was obtained (unpublished). Bitterman (**5**) in a study of transfer of decrement from one ocular task to another, apparently had no difficulty in setting up the conditions of a number checking test to produce a definite decrement in 30 minutes. On the basis of present knowledge, therefore, the indication that there are only slight fatigue effects in reading verbal text cannot safely be extended to numerals.

Tentative generalizations suggested by the work on numerals might be restated and amplified as follows:

1. So far as the material permits, tendencies to establish integrative habits and to compensate for unfavorable conditions will operate in the reading of numerals just as they do in the reading of verbal material.

2. With numerals as with verbal text, variables will interact to produce not easily predictable results.

3. Differences between the reading of numerals and of verbal text include: the nature of the psychological task, the relative importance of the different types of cues, and probably the ocular mechanics and resulting fatigue.

4. Because of greater dependence on geometrical details of form, the reading of numerals should be more amenable to analysis in terms of visual receptor theory.

Tables, Graphs, Scales, and Formulae

Washburne (**55**) studied various graphical devices and concluded that pictographs are suitable for very simple material, bar graphs for somewhat more complex static comparisons, line graphs for dynamic concepts, and tabulated round numbers for specific amounts. Croxton and Stein (**20**) found that bar figures can be compared most accurately by the reader, circles and squares next, and cubes least accurately. Graham (**24**) has also analyzed illusory tendencies in judging various graphic representations.

Tables, graphs, and scales[4] were compared for speed and accuracy in use in a group of Air Force studies by Carter (**11, 12**) and by Connell (**14**). It was concluded that:

1. Tables are preferable if the steps are small enough to make interpolation unnecessary.

2. Scales are slightly better than graphs if no interpolation is required.

3. When interpolation is required, scales and graphs are about equally effective and are superior to tables.

4. The frequency of coordinate lines on a graph, and whether the entry is on the x or y axis, make little difference.

The reading of mathematical formulae was studied by Tinker (**45, 46, 47**). He found, among other things, that it is characterized by "a tendency toward analytical reading and a breakdown of ordinary reading habits," and that formula contour can be an aid to perception.

The use of these various media enlists a battery of functions beyond the simple dis-

[4] "Scales" as here used means two scales representing the related variables plotted on opposite sides of the same base line.

crimination of digits. These include such functions as visual exploration, retention, perceptual judgment, and sometimes mental arithmetic. The above studies indicate, nevertheless, that these media for presenting information yield to experimental evaluation.

Maps and Charts[5]

Maps and charts are ways of representing selected characteristics of regions on plane surfaces. The term "map" is the more inclusive. "Charts" can be thought of as maps which give quantitative information for technical use, especially in navigation. In contrast to printed text materials which have been extensively investigated, maps and charts constitute an area in which the psychological problems have not even been systematically defined. A tentative classification of the functions of such instruments from the psychological point of view might be proposed as follows:

One function is the pictorial representation of an area so that it can be recognized when seen. The engraved coastal charts of the Coast and Geodetic Survey from the period when engraving was in its prime are examples of this type. Certain aeronautical charts, particularly the new experimental flight charts, serve the same purpose.

The second function is the provision of a symbolic aid for the establishing of a conceptual system. The conceptual system often takes the form of a more or less accurate mental picture from some imaginary point of view. The typical atlas map showing political boundaries presents such a picture in the most direct form. Contour lines, from which one skilled in their use can construct a picture of the topography, constitute a more indirect device for the same purpose. It is sometimes necessary to convey a system of relations in three-dimensional space between invisible elements. Wind charts and charts of such undersea conditions as temperature, currents, and salinity are examples. To speculate a little, the mental representation may be thought of as being in either visual, verbal, or other terms, depending on the individual. With less tangible phenomena such as culture traits the conceptualization is less likely to be visual.

A third function, the distinctive property of charts, is to present data for technical use.

The same instrument will often serve two or more purposes. Contour lines which give the geologist a conception of the land formation also give the highway engineer quantitative data about slope. An aeronautical chart may present a picture of the terrain and also data for navigation by radio. On the other hand, one function may conflict with another. A familiar example is the Mercator projection, whose rectilinear representation of compass direction is important for navigation, but whose distortion of land area is misleading in political maps.

The human capacities required in the use of maps would appear to include the more obvious visual discriminations, the ability to interpret several varieties of arbitrary symbols, the ability to structure perceptual fields in particular ways, and perhaps the development of complex imagery. The intelligence and education of those using the instruments is often important. Thus, the psychological processes in map reading are much broader than either the analytic discrimination of letters and digits or the stereotyped flow of ordinary reading.

In some respects the psychological problems that arise in the use of maps and charts are related to the problems of production. In addition to the techniques of printing and typography, such processes are involved as hand- and photoengraving, reproduction in color, the stereographic interpretation of airplane photographs, and the devising of new systems of projection. Creative as well as technological skills are demanded. Such skills are required of penmen in the shading of slopes by short parallel lines of graduated thickness known as hachuring. The task of

[5] Helpful background references in this field are Deetz (**21**), Greenwood (**25**), and Raisz (**36**).

compiling the material to be shown from a diversity of sources which have to be separately evaluated, and that of generalizing such features as contour lines and soundings to the best advantage when the scale is being reduced, require a high degree of professional judgment. Artistic skill is often called into play in the final arrangement of details to effect a pleasing appearance and to reflect correctly the relative importance of the various features in relation to the needs of the users.

A further complication in map making stems from the rapid changes in technology of recent years. Changes have been necessitated by demands for increased precision, for faster corrections and revisions, and for larger numbers of copies requiring long printing runs. They have been made possible by new techniques and materials such as photoaluminography and transparent plastic sheets. These developments have sometimes modified the perceptual effects produced. Lettering, for example, was formerly hand engraved, in styles which had been designed for the purpose. The current trend is to print lettering from type. As the preparation of map plates sometimes involves photographic reduction, a given type may ultimately appear in a size for which it was not originally designed. The photography sometimes results also in a loss of sharpness in internal corners. Legibility may suffer as a consequence. Another change is the decreased use of the stippled effect known as sanding for showing graduations of depth in coastal waters, with increased dependence on contour lines and tinting. This change saves time in corrections and revisions, but the new perceptual effect is quite different from the old.

In addition to technological changes, new circumstances of use may create new problems. Aviation has brought a need for flight charts with easily recognizable topography. Arctic flight has forced a consideration of the type of projections most suitable for polar regions. Red lighting which is designed to protect dark adaptation changes the contrast effects produced by tinting, and the low brightness increases the difficulty of discriminating fine print. Expansion of submarine activity has increased the need for effective portrayal of undersea conditions in three dimensions.

Complex and changing circumstances increase the cartographer's difficulty in keeping in rapport with operational needs. A minor example can be cited. There is an organized and efficient interchange of information between the makers and the users of nautical charts on harbor and coastal conditions, but the coordination seems to be less close on details of working arrangements. Submarine officers have pointed out that small table surfaces make it necessary to fold the charts, which hides the marginal scales and increases the difficulty of plotting. Repeating the scales at intervals on grid lines should solve the problem.

Returning to the psychological aspects of map and chart reading, we can ask what inferences seem indicated by the work on printed materials discussed above. One inference is that, with maps as with print, unpredictable interactions would necsesitate caution in extending experimental conclusions beyond their original context. With this limitation in mind, we can expect that the experimental approach would provide as valuable information in the one area as in the other.

Turning to more specific problems, it seems safe to consider that brightness contrast within the materials to be discriminated is a general principle of importance for visual work. This highlights the desirability of more information about the reading of symbols under low contrast as is often necessary with maps and charts.

A somewhat related proposition that appears generally valid is that the acuity and discrimination mechanisms are of increased importance as conditions approach thresholds, e.g., in type size or brightness. This directs attention to details of type form.

Type Forms

The work on printed materials indicates that when kinds of type differ in legibility, the difference can usually be understood in terms of such factors as size, boldness, internal and external spacing, and simplicity of form. On political maps especially, but to some extent also on charts, some names have to appear in small type. In view of the use of faint night lighting to protect dark adaptation, it might be profitable to design a type for maximum discrimination in terms of visual receptor capacities. Mr. S. W. Boggs, geographer of the State Department, who has analyzed the problems of map typography in detail, suggests that internal corners of the letters be exaggerated in the originals, as a protection against photographic and visual filling in.[6]

Type forms are commonly used to distinguish different classes of geographic features, both forms and sizes to distinguish sizes of towns. A study of easily distinguishable forms within the limits of good legibility would be worth while.

The legibility work indicates that lowercase words are more readable than capitals at reading distance. The relative merits of the two should be tested in the map situation.

Symbols

Maps and charts require a large number of standardized symbols besides letters and digits. Lists are readily available (**36, 60**). Some are isolated characters, others apply to areas and indicate physiographic features. The tradition is to make a symbol suggestive of the thing it represents, but many symbols have necessarily become conventionalized, and in some cases the differentiating details are minor. The work on letters and digits suggests that a search for areas of confusion among other symbols should prove worth while.

In relatively new areas, such as in the air and undersea, symbols have to be found to represent new concepts. It would be desirable under these circumstances to determine experimentally which of several proposed symbols suggests, on first impact to the largest number of people, the thing it is supposed to represent.

Color and Night Lighting

Among the problems of map and chart design for which the literature on printed materials affords little help is that of color. Certain conventions of color usage are sanctioned by wide acceptance: e.g., blue for water, green for vegetation, and red or black for cultural features. Other applications of color, as for elevation tinting, are still to some extent in the trial-and-error stage. Experimental pre-tests with colored materials are necessarily expensive, but probably less so than unsuccessful large-scale trials.

Night lighting creates special problems: first, that of getting enough light to make the charts readable without ruining dark adaptation, and second, the distortion of color values if the source differs in composition from daylight. The all-purpose maps on fluorescent paper, tonal-keyed for red light, are designed to solve the first of these problems. It has been reported that, at brightnesses which make the maps readable at lap distance, dark adaptation is impaired 0.6 log unit by a greenish-yellow fluorescent map surface and 0.3 log unit by a red surface, and that night acuity is impaired seriously by neither (**1, 34**). A current need seems to be for a stable material fluorescing red. (For data on several fluorescent materials, see Chapanis, **13**.)

Under red floodlighting, red symbols (e.g., for buoys) are not visible. Magenta, which shows dark under red light, has been substituted. Hue differences are, of course, lost, and brightness relations are changed. A thorough study of possible color arrangements, in relation to the spectral properties of the eye, available materials, and several light sources, is in order.

[6] Personal communication

Perceptual Patterns

Many of the creative skills of cartography have been applied to the task of giving valid perceptual representation to such features as coastal regions, mountains, slopes, and under-water contours by such devices as engraving, hachuring, sanding, tinting, and shading. In recent years the demands for faster production, economy, and more usable quantification of data have tended to shift the emphasis away from such pictorial characteristics. At the same time, aviation has tended to shift it back by demanding flight charts with easily recognizable landmarks. It is sometimes necessary, as, for example, in atlas maps and some aeronautical charts, to present so much information that it is difficult to select the desired items visually. An attempt to formulate in general terms the principles of perception relevant to these problems, drawing on both the experience of cartographers and the academic literature, would be profitable.

Unfavorable Conditions, Work Output, and Fatigue

In the routine of navigation, work with charts is likely to be alternated with other visual tasks over considerable periods of time. For reading under non-optimal conditions such as red lighting and vibration, the legibility data now available may not offer a valid index of the level of reader performance which can be expected in terms of speed, accuracy, and resistance to fatigue. It is therefore important that the literature reviewed above on visual performance under good conditions be supplemented by more systematic studies of those less favorable conditions, and that the problem of visual fatigue be investigated further.

Research Suggestions

A few of the broader areas for possible research suggested by this survey are recapitulated in the following list.

1. Since general theory serves the ends of economy in the long run, basic research in this field should be encouraged.

2. A clarification of the general problem of visual fatigue and energy cost on either the theoretical or the empirical level is to be desired. (Bitterman, Ryan, and Cottrell (**7**) are currently investigating the possibilities of the muscle-tension criterion in this connection.)

3. Applying the principles of visual acuity and discrimination to the designing of a type face of maximum perceptibility under near-threshold conditions might pay returns.

4. More extensive studies of the use of printed materials and charts under adverse conditions, when performance is measured as a function of several variables, would fill a noticeable gap in present knowledge.

5. Recent developments have created a need for more information about the relative effectiveness of different color arrangements in maps and charts, especially in relation to different light sources.

6. As a somewhat more speculative suggestion, an exploration of the principles of perception in relation to tables, graphs, maps, and charts might prove of interest.

Summary

Representative research on the factors of typography, illumination, and fatigue in normal reading has been reviewed, especially with reference to the more controversial methodological problems. Work on perception of type at a distance and on the legibility of numerals has been examined in relation to the factors in normal reading and to the characteristics of the visual receptors. Studies of tables and graphs have been summarized, problems of the use of maps and charts have been surveyed, and possible areas of further research in relation to printed materials, maps and charts have been suggested.

References

1. Armstrong, R. C. *Aeronautical charts prepared for use in daylight, under white, ultraviolet, red and amber lights.* USAF Air

Materiel Command, Memorandum Report No. ENG-49-698-18, Oct. 6, 1943.

2. Atkins, E. W. The efficiency of the eye under different intensities of illumination. *J. comp. Psychol.*, 1927, **7,** 1–37.
3. Berger, C. Stroke-width, form, and horizontal spacing of numerals as determinants of the threshold of recognition. I and II. *J. appl. Psychol.*, 1944, **28,** 208–231, 336–346.
4. Bitterman, M. E. Heart rate and frequency of blinking as indices of visual efficiency. *J. exp. Psychol.*, 1945, **35,** 279–292.
5. Bitterman, M. E. Transfer of decrement in ocular tasks. *Amer. J. Psychol.*, 1946, **59,** 422–438.
6. Bitterman, M. E. Lighting and visual efficiency: The present status of research. *Trans. Illum. Engng. Soc.*, 1948, **43,** 906–922.
7. Bitterman, M. E., Ryan, T. A., & Cottrell, C. L. Muscular tension as an index of visual efficiency: A progress report. *Trans. Illum. Engng. Soc.*, 1948, **43,** 1074–1081.
8. Bitterman, M. E., & Soloway, Esta. Frequency of blinking as a measure of visual efficiency: Some methodological considerations. *Amer. J. Psychol.*, 1946, **59,** 676–681.
9. Bitterman, M. E., & Soloway, Esta. The relation between frequency of blinking and effort expended in mental work. *J. exp. Psychol.*, 1946, **36,** 134–136.
10. Carmichael, L., & Dearborn, W. F. *Reading and visual fatigue.* Boston: Houghton Mifflin Co., 1947.
11. Carter, L. F. *The relative effectiveness of presenting numerical data by the use of tables and graphs.* USAF Air Materiel Command, Memorandum Report No. TSEAA-694-1, April 17, 1946.
12. Carter, L. F. *A study of the best design of tables and graphs used for presenting numerical data.* USAF Air Materiel Command, Memorandum Report No. TSEAA-694-1C, Sept. 24, 1946.
13. Chapanis, A. *Luminescent materials.* USAF Air Materiel Command, Memorandum Report No. EXP-M-49-695-12F, Nov. 4, 1942.
14. Connell, Shirley C. *The relative effectiveness of presenting numerical data by the use of scales and graphs.* USAF Air Materiel Command, Memorandum Report No. TSEAA-694-1M, Dec. 22, 1947.
15. Craik, K. J. W. *Instrument lighting for night use.* Flying Personnel Research Committee, Report No. 342, Great Britain.
16. Crook, M. N., Hoffman, A. C., Kennedy, J. L., Wessell, N. Y., & Wulfeck, J. W. *Further studies of the effect of vibration and other factors on legibility of numerals.* USAF Air Materiel Command, Memorandum Report No. TSEAA-694-1K, Oct. 21, 1947.
17. Crook, M. N., Hoffman, A. C., Kennedy, J. L., Wessell, N. Y., & Wulfeck, J. W. *Preliminary studies of the effect of vibration, illumination and type size on legibility of numerals.* USAF Air Materiel Command, Memorandum Report No. TSEAA-694-1F, June 23, 1947.
18. Crook, M. N., Harker, G. S., Hoffman, A. C., & Kennedy, J. L. *Studies of the effect of typographical spacing on the legibility of numerals under vibration.* USAF Air Materiel Command, Memorandum Report No. MCREXD-694-1Q, Dec. 23, 1948.
19. Crook, M. N., Hoffman, A. C., Kennedy, J. L., *et al.* (work in progress under contract with the USAF Air Materiel Command)
20. Croxton, F. E., & Stein, H. Graphic comparisons by bars, squares, circles, and cubes. *J. Amer. statist. Ass.*, 1932, **27,** 54–60.
21. Deetz, C. H. *Cartography: A review and guide for the construction and use of maps and charts.* Washington, D. C.: U. S. Gov't. Printing Office, 1943.
22. Dunlap, K. Report Highway Res. Bd., Nat. Res. Council, Division Office, 1932, App. E., p. 3, article 4.
23. Gleason, J. G. *The design of numerals for use in counter-type instruments: A review of the literature.* USN Office of Naval Research, Report No. 166-I-39, Dec. 20, 1947.
24. Graham, J. L. Observational and mathematical differences in bar graph and volume situations. *Psychol. Bull.*, 1936, **33,** 743. (Abst.)
25. Greenhood, D. *Down to earth: Mapping for everybody.* New York: Holiday House, 1944.
26. Hoffman, A. C. Eye-movements during prolonged reading. *J. exp. Psychol.*, 1946, **36,** 95–118.
27. Holmes, Grace. The relative legibility of black print and white print. *J. appl. Psychol.*, 1931, **15,** 248–251.
28. Keil, F. C. *Use of red light to facilitate dark adaptation.* USAF Air University School of Aviation Medicine, Project No. 118, Report No. 1, Feb. 19, 1943.
29. Luckiesh, M., & Moss, F. K. *The science of seeing.* New York: D. Van Nostrand, 1937.
30. Luckiesh, M., & Moss, F. K. *Reading as a visual task.* New York: D. Van Nostrand, 1942.
31. McFarland, R. A., Holway, A. H., & Hurvich, L. M. *Studies of visual fatigue.* Graduate School of Business Administration, Harvard University, 1942.

32. McFarland, R. A., Knehr, C. A., & Berens, C. Metabolism and pulse rate as related to reading under high and low levels of illumination. *J. exp. Psychol.*, 1939, **25**, 65–75.

33. Paterson, D. G., & Tinker, M. A. *How to make type readable.* New York: Harper and Bros., 1940.

34. Pinson, E. A. *Maps, charts, and photographs, fluorescent and tonal keyed for red light.* USAF Air Materiel Command, Memorandum Report No. EXP-M-49-695-120, Sept. 17, 1942.

35. Preston, Katherine, Schwankl, H. P., & Tinker, M. A. The effect of variation in color of print and background on legibility. *J. gen. Psychol.*, 1932, **6**, 459–461.

36. Raisz, E. *General cartography.* New York: McGraw-Hill, 1938.

37. Rebert, G. N. A laboratory study of the reading of familiar numerals. *J. educ. Psychol.*, 1932, **23**, 35–45.

38. Roethlein, Barbara E. The relative legibility of different faces of printing types. *Amer. J. Psychol.*, 1912, **23**, 1–36.

39. Rose, Florence C., & Rostas, S. M. The effect of illumination on reading rate and comprehension of college students. *J. educ. Psychol.*, 1946, **37**, 279–292.

40. Sanford, E. C. The relative legibility of the small letters. *Amer. J. Psychol.*, 1887, **1**, 402–435.

41. Stanton, F. N., & Burtt, H. E. Influence of surface and tint of paper on speed of reading. *J. appl. Psychol.*, 1935, **19**, 683–693.

42. Sumner, F. C. Influence of color on legibility of copy. *J. appl. Psychol.*, 1932, **16**, 201–204.

43. Taylor, Cornelia D. The relative legibility of black and white print. *J. educ. Psychol.*, 1934, **25**, 561–578.

44. Tinker, M. A. The relative legibility of the letters, the digits, and of certain mathematical signs. *J. gen. Psychol.*, 1928, **1**, 472–495.

45. Tinker, M. A. Reading reactions for mathematical formulae. *J. exp. Psychol.*, 1926, **9**, 444–467.

46. Tinker, M. A. How formulae are read. *Amer. J. Psychol.*, 1928, **40**, 476–483.

47. Tinker, M. A. A photographic study of eye-movements in reading formulae. *Genet. Psychol. Monogr.*, 1928, **3**, No. 2, 65–182.

48. Tinker, M. A. The relative legibility of modern and old style numerals. *J. exp. Psychol.*, 1930, **13**, 453–461.

49. Tinker, M. A. The influence of form of type on the perception of words. *J. appl. Psychol.*, 1932, **16**, 167–174.

50. Tinker, M. A. Effect of illumination intensities upon speed of perception and upon fatigue in reading. *J. educ. Psychol.*, 1939, **30**, 561–571.

51. Tinker, M. A. Illumination intensities for reading newspaper type. *J. educ.* Psychol., 1943, **34**, 247–250.

52. Tinker, M. A. Reliability of blinking frequency employed as a measure of readability. *J. exp. Psychol.*, 1945, **35**, 418–424.

53. Tinker, M. A. Validity of frequency of blinking as a criterion of readability. *J. exp. Psychol.*, 1946, **36**, 453–460.

54. Tinker, M. A. Illumination standards for effective and easy seeing. *Psychol. Bull.*, 1947, **44**, 435–450.

55. Washburne, J. N. An experimental study of various graphic, tabular, and textual methods of presenting quantitative material. *J. educ. Psychol.*, 1927, **18**, 361–376, 465–476.

56. Webster, Helen A., & Tinker, M. A. The influence of paper surface on the perceptibility of print. *J. appl. Psychol.*, 1935, **19**, 145–147.

57. Webster, Helen A., & Tinker, M. A. The influence of type face on the legibility of print. *J. appl. Psychol.*, 1935, **19**, 43–52.

58. Weiss, A. P. The focal variator. *J. exp. Psychol.*, 1917, **2**, 106–113.

59. Young, Katherine D. *Legibility of printed materials.* USAF Air Materiel Command, Memorandum Report No. TSEAA-8-694-1A, June 10, 1946.

60. *Standard symbols adopted by The Board of Surveys and Maps, United States of America.* Washington, D. C.: U. S. Geological Survey.

Chapter 3
THE DESIGN AND USE OF INSTRUMENTS[1]

WILLIAM E. KAPPAUF[2]

Princeton University

Introduction

Instruments are of many types and serve many purposes. For each type of instrument the problem of the design of its display is somewhat different, being defined in terms of the way or ways in which information is read or taken from the indicator and in terms of the varied conditions under which it will be used.

There are, in general, two kinds of instrument reading operation: one which proposes to be accurate and obtain a quantitative, numerical value from the indicator; the other, a monitoring or checking type of operation. Readings by either form of operation should conform with stated tolerances or operating requirements. How well these tolerances are met depends on the restrictive nature of the conditions of instrument use and on the adequacy with which display design variables have been manipulated to take particular account of the operator's characteristics.

In submarines, two considerably different sets of instrument reading conditions are recognized. Some instruments must be illuminated in such a way that the operator and his neighbors will remain as well dark-adapted as possible for night observation. Other instruments are used at stations where operator dark-adaptation is not a consideration and so these indicators and devices may be well illuminated. Level of illumination is, therefore, a factor which, in any thorough program of research on instrument displays, must be studied in relation to all other design variables. Currently available data on the effects of instrument illumination will be presented in a special section below. This will precede a discussion of the design of instruments for specific reading needs, but will follow a treatment of some general matters of instrument marking and numbering.

Instrument Marking

Instrument-marking problems will include, for purposes of the present discussion, all the more general aspects of the design of the instrument face. In particular, these are: contrast between the markings and the scale face, the length and thickness of scale graduation marks, the shape, size, boldness, and spacing of instrument numbers, and the labeling of the instrument. One does not proceed very far in the experimental study of these factors before discovering that they involve critical interactions of the sort already remarked on in Chapter 2.

Contrast between the Scale Markings and the Scale Face

When dark adaptation of the operator is important, it is appropriate to use instruments with white markings on black backgrounds. Adaptation level varies, in general, with the proportion of the visual field

[1] This chapter is based on portions of a review prepared by the author for the Air Forces, Air Materiel Command, Wright Field, Dayton, Ohio, and entitled, "Studies pertaining to the design and use of visual displays for aircraft instruments, computers, maps, charts, and tables: A review of the literature," USAF Technical Report 5765, April 1949.

[2] The author is indebted to William M. Smith, who, under an Air Force research contract with Princeton University, assisted in the preparation of the bibliography and the gathering of reference material which forms the basis of this chapter. (See **71.**)

which is bright and is protected by an instrument panel in white on black design (**49**). There has, however, been no systematic study to discover whether this contrast arrangement results in losses of instrument reading speed or accuracy when compared with instruments in black on white design. The one report which does refer directly to this contrast problem in relation to instrument reading *per se* provides no detailed data. It merely offers a summary statement that at high or daylight levels of illumination white on black dials demand a 25% longer reading time than black on white (**89**). This observation needs checking for both high and low illumination levels.

It was shown in Chapter 1 that visual discrimination improves as the degree of contrast between marking and background increases. Apparently the only exception to this generalization in practice is met when markings of low brightness are seen against a background of essentially "zero" brightness. Such a situation is met in the use of black dials with self-luminous or fluorescent markings. Operators consistently remark on the difficulty they have in accommodating on these markings which are described as scintillating and jumping. This difficulty is not experienced when visible red is used for instrument illumination, for under such light the dial background is visible. Thus, in order to facilitate accommodation, it is desirable to have less than maximum contrast in instrument displays when used under low illumination.

Thickness and Minimum Length of Scale Graduation Marks

Choice of the thickness for scale graduation marks, when the space between marks is fixed, involves a compromise between the demands for mark legibility and the demands of precision of interpolation between marks. Thus, although bold markings must be used on instruments intended for use in low illumination, stroke boldness in itself may lead to interpolation errors and lack of reading precision in the neighborhood of the mark. To date, investigations of graduation-mark dimensions have considered scale reading at high illumination levels only.

Maier (**58**), concerned with the specific non-interpolation problem of reading a stopwatch to the nearest fifth-second mark, conducted an experiment to determine what the optimum mark thickness should be. In eight different scales, mark thickness was varied from 14% to 53% of the scale range occupied by each fifth of a second. Maier's subjects, all of whom had had some training in instrument use, read the different scales in projected tachistoscopic exposures of less than 0.4 seconds. Errors were fewest for the scales on which the graduation marks occupied about 25% of the graduation interval. Appreciating the possible limitations of his tachistoscopic technique, however, Maier cautions that this result might not be confirmed for other exposure periods.

The length of scale-graduation marks was considered by Schulz (**69**). In a not too rigorous, apparently casual way, he recorded reading times for a variety of scaled instruments. When the data were analyzed, Schulz concluded that reading was best when the ratio of length of the shortest scale marks to mark separation was in the order of 1:1 or 2:1. Larger ratios, some as high as 6:1, were considered less favorable to good reading. Unfortunately, Schulz presents no details of his procedure, no information about his subject population, and no specific performance data to support his generalization. An experiment in this area still needs doing.

Instrument Numbering

Numeral Contours and Stroke Thickness. Research on the subject of the design of numerals for use in printed materials has been reviewed in Chapter 2. The studies there referred to, Tinker (**75, 76**) and Dunlap (**21**), and other papers by His Majesty's Stationery Office (**94**) and by Lauer (**46**), drew attention to the importance for numeral legibility of such design factors as stroke width and any variations there, of, the area of white space included within figure outlines, and the spacing of grouped digits. It remained, however, for Berger (**7, 8**) to make the first clearly systematic study of the whole problem. Berger's experimental variables included stroke width, numeral shape, spacing and lighting. His measure of legibility was the maximum distance at which digits of a given size could be distinguished.

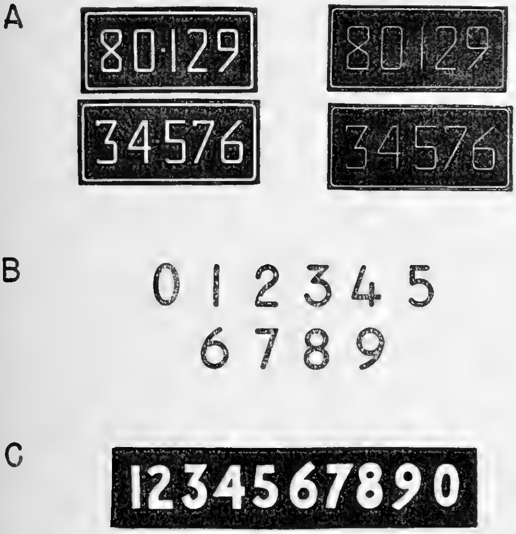

Fig. 1. Three sets of digit forms
A. Berger's digits (**7, 8**) showing optimum stroke thicknesses for white on black displays seen by reflected light and by transmitted light
B. U.S. Army-Navy standard instrument dial numbers (**91**)
C. Craik's digits (**19**) intended for white on black displays

produced by transmitted light. These tests showed that optimum stroke width is a function of the specific illumination and contrast conditions. Expressed as a fraction of numeral height, the optimum widths found by Berger were as follows: 1/13th letter height for diffusely lit white figures on a black background; 1/8th for diffusely lit black figures on a white background; and 1/40th for bright figures presented by transmitted light. His results for the diffusely lit white on black and black on white figures are

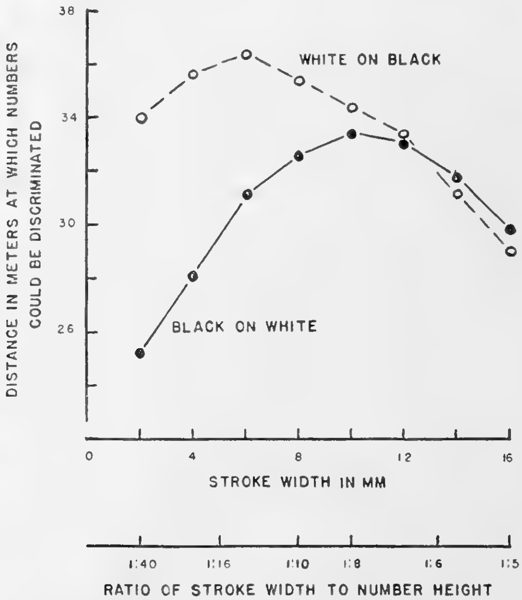

Fig. 2. Influence of ratio of stroke to height and of direction of contrast upon the legibility of Berger's digits. Data for two subjects who observed approaching 80-mm. numerals until they could be correctly read (**7**).

The forms which Berger arrived at are shown with two comparison sets of numerals in Fig. 1.

Berger began by studying stroke width. He proposed to find what width would permit discrimination between the numerals 2, 5, and 8 at the greatest distance. Experiments were conducted out of doors under three conditions of lighting and contrast: (1) with diffuse light, black figures on a white ground; (2) with diffuse light, white figures on a dark ground; and (3) with bright figures

shown in Fig. 2. These data, based on the observations of two subjects, show neatly the nature of the interaction between contrast direction and stroke width. Each type of figure had its own optimum stroke thickness. Note that the best white on black figures were readable at greater distances than the best black on white ones.

When Berger investigated digit shape, he did so by varying the form of each digit until its legibility was equivalent to that of the number 8 used as standard. The letter forms which he finally developed are quite different from those in general use on instruments or in printed copy. Before being adopted for general use, they must, of course, be evaluated for other sizes and reading distances, and by criteria other than the one which Berger used. Important, however, for the present discussion is the fact that Berger's numeral shapes were empirically determined. It was on the basis of experimental evidence, and not just a series of "hunches," that he decided (1) to make the angle at the base of the 2, in the 7, and in the tails of the 6 and 9 as near to 45 degrees as possible, (2) to close the angle at the top of the 4, (3) to run the bar on the 5 over the entire number, and (4) to make the center point on the 3 midway fall between right and left borders of the figure. Some agreement with Berger's designs, particularly for the numbers 2, 4 and 9 is found in studies of time-table digits which Fishenden (**26**) reviews very hastily.

The numbering of most service instruments in this country today follows a modified Le Roy lettering style, like that shown in Part B of Fig. 1. Changes introduced in this standard style from time to time appear to have been made principally on a priori grounds (**93**). Currently, the recommended ratio of stroke thickness to numeral height in this series is 1:6.

British work on numeral designs seems to have been limited to that of Craik (**19**). He determined the legibility of instrument numbering by observing the speed with which trained subjects read numbers at different brightness levels. He summarized his results by saying that at any given digit height, legibility was greatest when the white parts of the figures were equal in breadth to the mean width of the enclosed black parts. This can be interpreted as signifying a ratio of stroke width to digit height of 1:5. Craik's number series is shown in Part C of Fig. 1. Confusion between the 6, 8, and 9 was said to be less likely when the digits were of the form shown.

Stroke thickness as such as been the subject of several other studies. An early series of tests by Paterson, Walsh, and Higgins (**60**) on the legibility of self-luminous numerals pointed, as did Berger's later data on luminous numerals, to the desirability of low stroke-width to digit-height ratios. More recently, Loucks has compared the readability of instruments marked with numbers of different degrees of boldness. In the first experiment (**51**), he compared numbers 0.19 inches high with stroke widths of about 1/6th and 1/4th letter height. Under ultra-violet illumination, at what is described as a relatively intense level, dials numbered with these two different sets of digits were read in tachistoscopic exposure with equal accuracy. Under a higher tungsten illumination, however, the dials with the 1:6 numerals were read with significantly fewer errors. In a companion study of numeral markings on climb and dive indicators (**52**), comparable results were obtained. In ultra-violet illumination dials with 1:16 numerals were read less accurately than dials with 1:8 or 1:5 numerals. In tungsten illumination, however, the 1:16 and 1:8 numerals were equally satisfactory and superior to the 1:5. These data are consistent in suggesting that the optimum boldness of white on black figures decreases as illumination intensity increases.

Digit Proportions. The problem of numeral proportions, the ratio of numeral width to height has received little, if any, experimental attention. Loucks (**51**), in discussing possible extensions of his research, suggests the need for legibility studies with numbers of proportions different from the ones used on his test scales. He states that "tall but narrow overall width figures . . . have been found to give maximum legibility in combination with minimal interference with night vision." No reference for the latter comment is given, however. Such a change in numeral proportions corresponds to the change made in going from normal to "condensed" type. In as much as evidence from some studies of typography indicates that certain condensed types, especially if bold, may be less legible than normal, the problem of instrument numeral proportions would seem to be worthy of further special study.

Berger studied digit proportions in one of his experiments (**7**) but only after he had arbitrarily chosen a width to height proportion of 42:80 for his standard numeral, the digit 8. He then determined what proportions each of the other digits should have in order to be equal in legibility to the 8. Except for the digit 1, all figures required proportions similar to that of the 8, namely about 1:2. The exact values of the figure width recommended by Berger are presented in Table I. It is interesting to note that these 1:2 proportioned digits of Berger depart from the 2:3 proportioned digits in the U.S. Army-Navy Aeronautical Design Standard (see Fig. 1) and that the departure is in the direction suggested by the foregoing quotation from Loucks.

TABLE I

Berger's Data Showing Optimum Width and Spacing of Digits Which Are 80 mm. High

Digit	Width	Spacing on right side	Spacing on left side
	mm.	*mm.*	*mm.*
0	39.3	12.3	12.3
1	6.0	11.1	11.1
2	39.2	8.5	9.9
3	40.4	7.5	8.9
4	48.0	10.2	10.9
5	40.5	9.1	8.5
6	41.2	7.7	8.1
7	43.0	5.5	3.5
8	43.0	8.9	8.9
9	41.2	8.1	7.7

Digit Spacing. The data which Berger collected on digit spacing (**7**) are also presented in Table I. The spacings given in the table are additive for any two adjacent figures: e.g., in the number 10, a space of 11.1 mm. should be allowed to the right of the 1 and another 12.3 mm. to the left of the 0, making an over all space 23.4 mm. between the two digits. These spacing values were determined principally by finding how far each digit had to be separated from zero in order for the combination to be legible.

In view of the repeated reference made here to Berger's results, it is necessary to remark that they represent only one approach to the digit legibility problem and that Berger himself warns against adopting his results unquestioningly for application in the design of displays other than license plates. Since all of his data were based upon the reading of distant stimulus cards, comparable investigations to check his results at normal reading distances are needed.

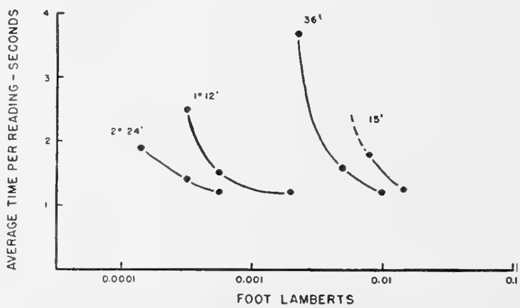

Fig. 3. Number-reading time as a function of brightness. Each curve is for a different size of number, being given in visual angle. Data from Craik (**19**) for two subjects.

Digit Size. The most pertinent data on the subject of instrument numeral size are those of Craik (**19**). Working with only two subjects, Craik determined reading time for numbers of four different sizes as a function of dial illumination. His number sizes are given as angular size. The plot of average reading time vs. brightness in foot lamberts is reproduced in Fig. 3. Craik would suggest the use of this chart, or similar ones based on more adequate data, for selecting instrument numeral size once illumination level and observation distance have been fixed for a given operating situation.

Instrument Labeling

Two principles have been applied in support of the suggestion that instrument names and labeling be cut to a minimum. The first is the rule of keeping markings at a minimum in order to preserve operator dark adaptation when this is important. The second is the rule evolved by Loucks (**50, 52**) that instrument faces should be kept as "clean" as possible. Neither discussion nor research on this labeling problem has been extensive, however. At the discussion level there is Craik's suggestion (**19**) that instrument titles be abbreviated to two or three letters. At the experimental level there is Loucks' finding (**52**) that when the words "Up" and "Down" are eliminated from the climb and dive indicator, accuracy of reading is actually improved.

Unfortunately, even when instruments are thought to be well labeled, misreading one instrument for another is very easy on a busy panel with many indicators. Confusions of this sort accounted for 7% of all the serious "pilot errors" reported in a recent survey by Fitts and Jones (**27**). As Fitts and Jones point out, there is need for the development of a practical system which will insure the easy and positive identification of each type of instrument.

From the foregoing discussion it is apparent that currently available research information on instrument marking is limited and that in too many cases the data reported are for very few subjects. Of particular interest for the immediate future will be an extension of the studies of the interactions between the variables of contrast direction, illumination level and stroke thickness for markings. Research papers which approach aspects of this problem and which should be consulted in the planning of future work include those of Berger (**6**), Byram (**13**), Ferree and Rand (**22, 23**), Fry and Cobb (**28**), Holmes (**39**), Taylor, (**74**), Walls (**81**), Weston (**84**), and Wilcox (**85, 86**).

Instrument Illumination

Color and Method of Illumination

In Chapter 1 evidence was presented to show that in work situations where operators must remain well dark adapted it is desirable to use red lighting. Ultra-violet lighting of fluorescent pigments does not disturb dark adaptation if kept at a minimum level for instrument reading (**10**), but since red fluorescing pigments are not available, an ultra-violet system is less safe than a red lighting system when the illuminating intensity happens to be set too high.

Spragg and Rock (**73**) in a recent study have shown that dial reading speed and accuracy are unaffected when filters are used over the light sources to cut out all but the red end of the spectrum. They compared the performance of subjects reading sets of twelve dials seen under light passed by each of four sharp cut-off filters: wavelengths 440 mμ and longer, 550 mμ and longer, 620 mμ and longer, and 640 mμ and longer. Observations were made at 0.1 and 0.01 foot lamberts. Further data seem to be needed to clarify the present records for reading under the 640 mμ filter, but up to 620 mμ there was no loss in reading speed or accuracy at either brightness level.

Of the several methods of instrument illumination which might be used, flood lighting and edge or rim lighting are the two deserving of most consideration. In general, flood lighting is preferred because it provides more even illumination. No specific tests have been reported, however, on visual performance under these two forms of lighting.

Intensity of Illumination

When really precise instrument scale reading is needed, high illumination intensities should be used. Craik (**19**) and Bartlett (**4, 5**) both point to the need for brightnesses above 10 foot lamberts on the basis of general visual data of the sort reviewed in Chapters 1 and 2. Crouch (**20**) reviews data on scale reading speed which are to the same effect. These are data of Ferree and Rand, presumably not published elsewhere. They show that the speed of reading a steel vernier rule with black markings increases through the brightness range of 20 to 200 foot lamberts. Correspondingly, Grether and Williams (**36**) found that dial reading speed increased when simulated day conditions, i.e., white floodlighting at 45 foot candles, were substituted for simulated night reading conditions, i.e., ultra-violet illumination. Loucks (**50, 51, 52, 53**) has data which suggest that, in tachistoscopic exposures, instruments which are more intensely illuminated are read with fewer errors.

Two sets of data have been collected for the specific purpose of determining the effect of low illumination intensities upon instrument reading performance. One was obtained in an experiment by Craik (**19**) in which several forms of instrument illumination were tested. The other is reported in a paper by Spragg and Rock (**72**) and describes the effect of intensity for white illumination only. The conditions of these two experiments were quite different. Spragg and Rock used a serial reading task wherein twelve 2.8 inch diameter dials were read in succession at a 28-inch reading distance. Craik, on the other hand, took stop watch times for single readings of a 3.75 inch diameter dial. In terms of scale length, Spragg and Rock's subjects were required to read to the nearest 0.09 inches, Craik's to the nearest 0.18 inches. The data of the two experiments are summarized in Figs. 4 and 5. Craik's results are shown for reading times only, Spragg and Rock's for reading times and reading errors.

It will be noted first that the time curves for reading under white light in the two experiments are not similarly located on the brightness scale. The difference in their location with reference to the brightness axis is to be accounted for at least in part in terms of the differences in dial size and in size of the figures and marks to be discriminated. Observation distance may also have been a factor. But, more significantly, the curves are of different shape. Craik's curves, as drawn, resemble curves obtained by Richtmyer and Howes (**65**) and others for the effect of illumination intensity upon the rate of reading printed material. That of Spragg and Rock, with a step and a second level portion, does not. At least two questions need answering. First, would Craik's curves have had a shape similar to Spragg and Rock's if he had used lower brightnesses

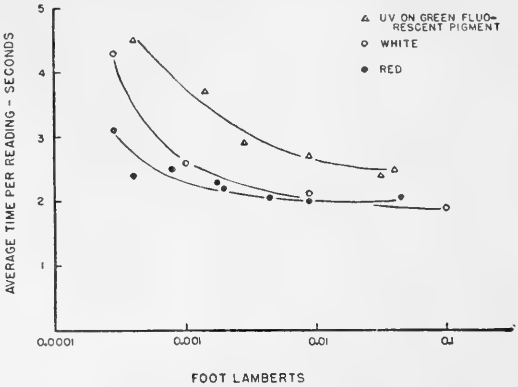

Fig. 4. Dial-reading time as a function of brightness for scales illuminated by white, red, and ultraviolet light. Data from Craik (**19**) for not more than six subjects reading under stop watch timing.

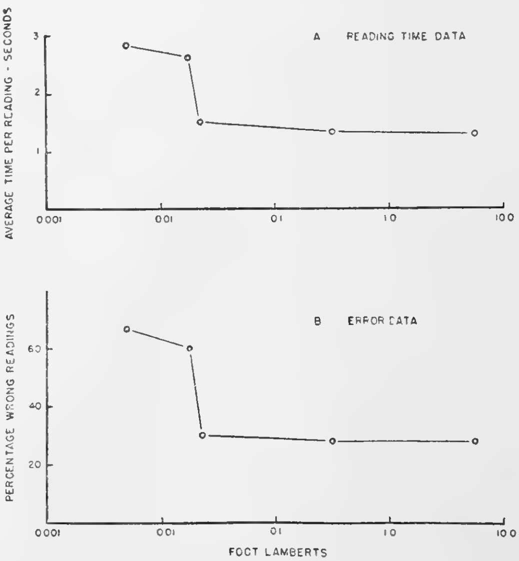

Fig. 5. Dial-reading performance as a function of brightness for scales illuminated by white light. Data from Spragg and Rock (**72**) for 20 subjects, each of whom made 100 readings at each brightness level under instructions to read as rapidly and as accurately as possible.

or levels which were nearer together in the region where the brightness effect begins? The variability of the observed points plotted for the several illuminants in Fig. 4 does in fact leave open the question as to whether Craik drew on his graph a best curve of best fit. Craik chose a simple function but a step-shaped curve would probably have fitted his data as well. And second, does the step in the Spragg and Rock data reflect some change in the nature of the reading performance? Did the subjects experience a change, for example, from reading under conditions of "good seeing" to reading under conditions where they could not discern the individual scale marks or see the numbers clearly, but where they could still infer the approximate value associated with a given pointer setting because of their familiarity with the scale and its numbering? The error records shown in Part B of Fig. 5 reveal that the subjects' accuracy changed as abruptly as their reading speed at the critical brightness level. This occurrence of corresponding error and speed changes suggests again the desirability of knowing more about performance in this region. A general reconciliation of these results with those of Craik will be welcomed. New studies may well be concerned also with the question of why the critical brightness levels in the two sets of results were so widely different.

It is of interest to observe the specific points plotted for red and white illumination in Craik's graph, Fig. 4. With the exception of readings at the lowest intensity of red light, reading under red and white light was equally rapid. These data are in line with the results of Spragg and Rock's filter study discussed above.

Reference may be made again at this point to Fig. 3 dealing with the effect of intensity of illumination on reading speed for instrument numbers of different size. One partial check on the adequacy of these data is found in the results of Lee and Finch (**49**), who determined what the brightness of red illuminated instrument numbers should be in order for them to be clearly legible. For numbers 18 minutes by 14 minutes in angular size, the required brightness was 0.009 foot lamberts. This value agrees satisfactorily with Craik's determinations for 15-minute numbers.

Instruments for Presenting Quantitative Data

Counters vs. Scales

For indicating precise numerical information, a counter type of instrument can probably not be surpassed. If such an instrument is well engineered, the reader has no interpolation problem. He simply reads the digits of the number as they appear in the counter aperture. Design problems reduce to those of making the numbers legible and spacing them properly.

Comparative data on the relative merits of scales and counters for the presentation of quantitative information can be found in Grether's study of altimeter designs (**31**). The counter was read faster and with fewer gross errors, of 1,000 feet or more, than the simplest scaled instrument tested, a straight vertical scale. (See Fig. 11 below.) Similar results were obtained by Chapanis (**14**). As the expression "direct reading counter" implies, a counter requires essentially no interpretation. This makes the counter a desirable indicator in situations where specific quantitative data are to be read.

There are situations, however, in which the counter is not the most satisfactory form of indicator. One of these is for the task of instrument setting or adjustment. An operator who is trying to come to a required setting using a counter sees nothing but a whirl of digits. Thus, the counter developed by Riggs for field artillery application (**9, 66**) and another by Chapanis for presenting radar scope bearing data (**15**) were recommended for use in reading bearing information but not in making bearing settings. For setting operations, dials or scales are by test more satisfactory.

Scaled instruments, of course, have other advantages peculiar to themselves which make them useful for many functions. They can often be set up in such a way as to be semi-pictorial. They can also be designed in such a way that makes for rapid check reading when precise quantitative data are not needed. Thus, the scaled instrument has received a great amount of research attention in order that its precision might be raised as much as possible and its limitations known more fully. These studies have dealt on the one hand with certain perceptual problems in scale reading, number preferences and the like, and on the other with the details of scale design and scale form.

Number Preferences and Errors of Interpolation

A person is said to exhibit a number preference, if in a number or scale reading situation, the frequency of his replies for each digit departs systematically from the distribution of digits presented. Thus, if in the number series read from a "randomizing machine," which over a long time period presents the digits 0, 1 . . . 9 with equal frequency, some digits occur considerably more often than others, then the person who recorded those data displays a number preference or bias. Such biases have been observed among persons using these machines to set up random number tables (**44**). They appear also in scale reading tasks where subjects are required to interpolate or round their readings (**1, 2, 18, 87**). It is the rare subject, for example, who uses all the final digits with equal frequency when reading to tenths of scale intervals. The most common pattern, if one dares to generalize from the not always consistent data, shows a massing of readings at the 0, 2, 5 and 8 positions. Scale reading errors are, of course, least frequent when the pointer or index is at a major division mark, and within divisions, are least frequent when the pointer is at the middle of the division range. The increase in reading errors when the pointer is set to values either side of middle (**42, 43, 63, 64, 67**) is in part a function of the aforementioned preferences or rounding tendencies.

The implication of these data for scale design is that subjects cannot be expected to read to tenths of divisions with complete accuracy. Their tendency, particularly apparent as graduation-mark separation becomes small, is to cluster readings around fourths. Quite probably the improvement of scale reading accuracy in situations where high precision is demanded requires the use of scales graduated by twos or units. Exactly how much improvement these graduation schemes effect is in need of investigation, for, as will be seen below, good comparative data on reading accuracy are available only for scales graduated by fives and by tens.

Details of Scale Design

At least three criteria have been applied in evaluating the scale-design variables to be considered in the following paragraphs. Which design permits the fastest reading? Which design is read with the fewest errors? Which design is read with the fewest large, and generally serious, errors?

Graduation-Mark Separation, Interval Values, and Scale Length. Research evidence has recently been accumulating which will assist designers in spacing graduation marks and in choosing interval values for scales on instruments which are to be read within stated accuracy limits. Such scale design, based on operator reading characteristics and upon stated performance standards, will mark a real advance over the relatively arbitrary scaling principles employed in the past.

The primary questions which have been raised by these recent experiments are: (1) What is the relative speed and accuracy of reading scales which are alike in terms of the value assigned to the smallest subdivision but which differ in the scale length devoted to each subdivision? (2) What is the relative speed and accuracy of reading scales which are alike in terms of the scale length devoted to a given number of scale units but which differ in the number of subdivisions and hence in the value of the subdivisions used? (3) Does over-all scale length or dial size interact with the foregoing effects?

For the design of scales which are to be read to the nearest unit, the only graduation schemes worth considering are graduations by units, twos, fives, or tens. Vernon (**79**) found it not difficult to show that scales graduated in other ways were read with more frequent errors.

Studies of the accuracy of reading scales graduated by tens but using different scale lengths per division are in agreement in indicating that the accuracy of reading to tenths of divisions increases as the distance between graduation marks increases to about 0.75 inches (Grether and Williams, **36**; Kappauf and Smith, **41**). The agreement among these results is the more encouraging in view of the differences in test conditions and the variation in stimulus materials which were used in the experiments.

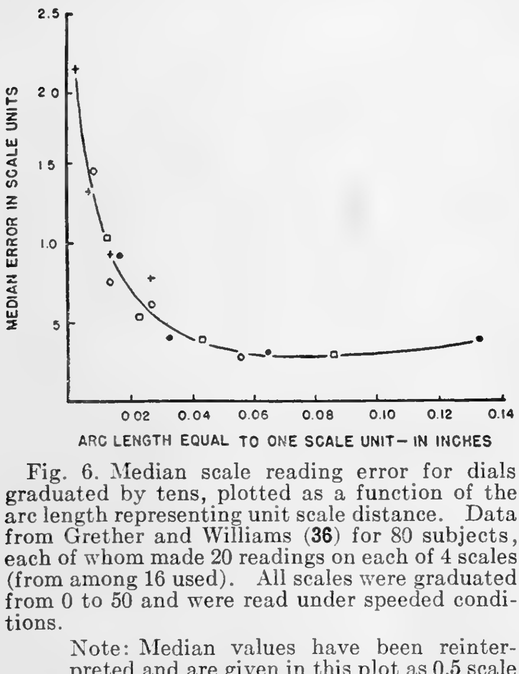

Fig. 6. Median scale reading error for dials graduated by tens, plotted as a function of the arc length representing unit scale distance. Data from Grether and Williams (**36**) for 80 subjects, each of whom made 20 readings on each of 4 scales (from among 16 used). All scales were graduated from 0 to 50 and were read under speeded conditions.

Note: Median values have been reinterpreted and are given in this plot as 0.5 scale units less than those given in the original report.

Grether and Williams used a speeded reading situation where response times to the presentation of single instruments were measured. All scales ranged from 0 to 50 but occupied different amounts of the circumference of four differently sized dials. Kappauf and Smith used a serial reading task, 12 dials all alike in design being read in sequence under accuracy instructions. The scales occupied the full circumference of every dial, some scales ranging from 0 to 100, others to values of 200, 400, and 600. The results of these two experiments are shown in Fig. 6 and in Part A of Fig. 7. Note that the curves reach low points, indicating small average errors for Grether and Williams or infrequent errors for Kappauf and Smith, as arc length equal to one scale unit approaches 0.05 to 0.1 inches, or as mark separation approaches the size 0.5 to 1.0 inches.

The accuracy of reading scales which require interpolation to fifths of intervals has been studied by Kappauf and Smith (**41**). Here it appears that when subjects are reading under accuracy instructions there is little reduction in the frequency of their reading errors when the arc length equal to one scale unit is increased beyond 0.02 or 0.04 inches. See Part B of Fig. 7.

The foregoing paragraphs can be summarized by saying that the evidence available at the present time suggests that for greatest accuracy in reading scales to units, the distance allocated to each scale unit should be between 0.05 and 0.1 inches on scales graduated by tens, and need not exceed 0.04 for scales graduated by fives. Comparable data for the reading of scales graduated by twos or units are not available.

On the converse problem of the accuracy of scale reading when unit scale distance remains fixed and graduation values change, the following data may be cited: For unit scale distances exceeding 0.01| inches, scales graduated by fives are read with fewer errors than scales graduated by tens (**40, 41**). This can be seen by comparing Parts A and B of Fig. 7. For the case of unit scale distance equal to 0.12 inches, graduation by ones is superior to graduation by fives for clock reading (Grether, **29**). Information on the effectiveness of graduation by twos is inadequate at present.

Discussion in the preceding paragraphs has treated scale reading accuracy as dependent on the linear separation of graduation marks or on unit scale distance and, for circular scales, as independent of dial diameter. The data of Williams and Grether which are presented in Fig. 6 represent readings on dials of four different diameters, from 1 to 4 inches. The data of desired ranges of values. Should the reading tolerances for any particular task be such that there is no need to employ the most effective unit scale distance or scheme of graduation, Figs. 6 and 7 may be used to suggest to what limits the graduation mark spacing may be reduced.

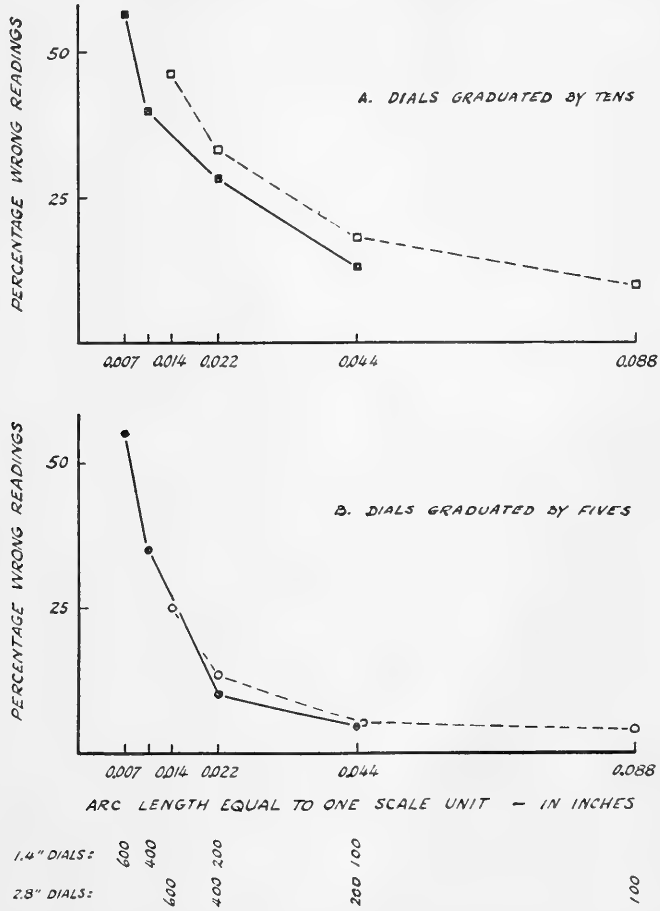

Fig. 7. Frequency of reading errors for dials graduated by tens and by fives. Data from Kappauf and Smith (**41**) for six subjects, each of whom made 30 readings for each type and size of dial under instructions to be as accurate as possible. Brightness level: six foot lamberts.

Kappauf and Smith in Fig. 7 are for two sizes of dial, 1.4 and 2.8 inches in diameter. The two figures suggest that the foregoing assumptions are not far from the truth. The subject is currently under further investigation, but for the present it seems reasonable to use the unit scale distance data cited above in designing scales to cover any

Kappauf and Smith (**41**) report the reading time data shown in Fig. 8. In their test situation, requiring the serial reading of 12 dials of identical design and for the full 360 degree scales which they used, reading time proved to be related primarily to the number of units represented on the scale and was relatively uninfluenced by dial size within the size limits studied. Thus, the data in Fig. 8 are plotted in terms of unit scale length in degrees. Note that the average reading times of these accuracy motivated subjects ranged from about two to five seconds.

Schemes of Varying or Accenting Gradution Marks. To make the graduation scheme easy to interpret, main division marks are sometimes lengthened, made bolder, marked with characteristic symbols or left incomplete or shorter than other marks at the scale circumference, so that the scale appears to have a break in it.

Maier (**58**), in the course of comparing reading accuracy for several commercially available stopwatch faces, tested some scales with and some scales without the "break" just referred to. His conclusion was that for scales which are to be read to the nearest mark, the scale break is undesirable.

Christensen (**16**) recently proposed a method of varying graduation mark length which may prove useful on special scales where readings are ambiguous or where errors of reversal, i.e., reading the scale in the wrong direction, occur frequently. Christensen's scheme, called staircase scaling, involves the use of progressively longer marks to designate each of the un-numbered graduations within a larger division. Mark length thus becomes a symbol for the numerical value represented by the mark. Sample scales in this form are shown in Fig. 9. In a preliminary test of reading performance on staircase scales and on conventionally graduated scales (**17**), one of two groups of subjects made significantly fewer reversal reading errors with the staircase. The second group of subjects, composed of pilots familiar with conventional scales, appeared to take some time to learn the use of the staircase, for only in the second half of the experiment did

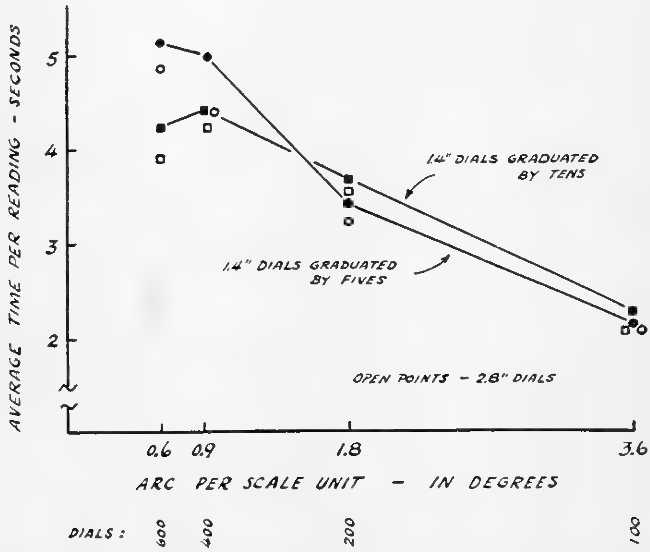

Fig. 8. Average time per reading for dials graduated by tens and by fives. Data from Kappauf and Smith (**41**) for six subjects, each of whom made 30 readings for each type and size of dial under instructions to be as accurate as possible. Brightness level: six foot lamberts.

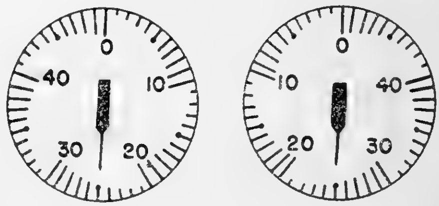

Fig. 9. Examples of the staircase scale. From Christensen (**17**)

they, like the first group, make significantly fewer reversal errors on the new scales. In the course of this experiment both clockwise and counterclockwise scales were used. Since Christensen did not analyze his data to observe the effect of the staircase on each direction of scale separately, it is possible that further useful information could be obtained from a re-analysis of his records. Further study of this method of scaling will be anticipated with interest.

Scheme for Numbering Graduation Marks. Some results obtained by Vernon (**79**) link scale numbering with the graduation problem. Wherever possible, it appears desirable to have different scales which are to be read in succession designed with the same numbering and the same graduation values. Vernon observes, however, that if graduation interval values must be different for scales which are to be read successively, less confusion is caused by using the same number series on the scales and varying the intervals between them than by keeping the number of divisions between numbers constant on all scales. This problem deserves reinvestigation because it is not clear to what extent Vernon's conclusion, from admittedly preliminary data, resulted from the exceptionally confusing graduation values used on some of her scales.

Without offering specific supporting data, Vernon advances the following principles or rules of scale numbering: (1) that numbers on the scale should not be so close together as to cause confusion, (2) that large numbers are to be avoided, two-digit numbers being better than three, (3) that decimals are to be avoided, (4) that numbering by tens (or tens multiples thereof) is most desirable, (5) that the maximum, and probably the optimum, number of divisions between scale numbers is ten. These rules appear to have been reasonably well recognized by scale designers in the past.

The only reported experimental inquiries into the question of the number of numbered graduations to have on dials which are actually in service use are those of Loucks (**50, 51, 52, 53**). Loucks studied various designs of tachometer, climb and dive indicator and manifold pressure dials. His procedure was to expose individual dials for short in-

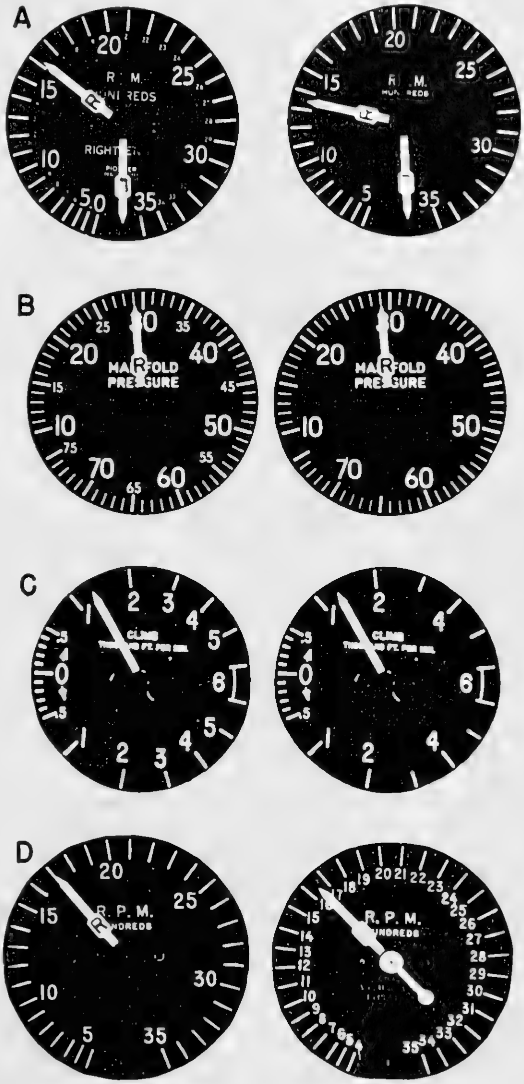

Fig. 10. Examples of pairs of dials tested by Loucks (**50, 53**)

tervals, usually 0.75 seconds. He used tungsten, red and ultra-violet illumination in various test runs. In general, he found that there was no disadvantage and sometimes an advantage to dropping some of the dial numbers. Fig. 10 compares the dial pairs for which the following results were obtained. On the tachometer dial he found that the small $\frac{1}{16}$-inch numbers marking every 100 rpm from 4 to 35 could be dropped, leaving only the seven larger numbers which marked each 500 rpm. Accuracy was unaffected. A similar elimination of seven small $\frac{1}{8}$-inch numbers from the manifold pressure dial, leaving just the seven large ones, left performance unchanged in tungsten light and improved reading accuracy in ultra-violet illumination. No loss attended the dropping of several numbers from the climb and dive indicator, even though the subjects in the experiment expressed preference for the fully numbered scale. When, as in one instance, results significantly favored the more completely numbered dial (see Part D of Fig. 10), they occurred in this direction only for an exposure of 1.5 seconds and not for 0.75 seconds where, indeed, the simpler numbering had a significant advantage under ultra-violet illumination. This difference of result for these two different exposure times could be interpreted as meaning that the amount of usable scale numbering may vary with the kind of reading being made; that for reading "at a glance," infrequent numbering with easily comprehended "round" values may be desirable, while for more careful and leisurely reading, frequent numbering may be an aid. A test of this conjecture with longer exposure times or in a "read at your own pace" situation is indicated.

No research appears to have been done on numbering arrangements for instruments used under low illumination.

Non-uniform Scaling. Instrument designers are sometimes forced to use scales with non-uniform graduations. Some instruments are mechanized in a way that pointer movement is non-linear. Others require readings in such widely different parts of the scale range that non-linear scaling is one answer to the problem of getting the entire scale into the available space. No data are yet available, however, on how reading accuracy is affected by scale non-linearity, or on the problem of how graduation and numbering changes should proceed along such scales if reading errors are to be minimized.

Broken or divided scales. The need for a long range of scale values commonly suggests the use of broken scales—that is, of a coarse scale coupled with one or more fine scales. This has been mechanized in a variety of ways: as a large dial with a small insert dial, as a pair of large dials, as a single dial with several pointers. But no matter what form the broken scale has taken, gross inaccuracies in making certain readings have always resulted. In field artillery these errors have taken the form of the much publicized "100 mil error" (**9, 88**). In aircraft operation they have occurred with serious frequency in altimeter reading as errors of 1,000 feet or 10,000 feet. At best, the broken scale is difficult to read because the operator has to combine several separate readings to get his final value, but the particular difficulty which produces large, systematic errors arises when the coarse scale registers just below but near a value which would imply a full circuit of the fine scale. On a coarse and fine scale combination, for example, where the fine scale registers units and tens up to 100 and the coarse scale registers hundreds, a setting of 397 is often read as 497 because the coarse scale pointer is read, not to the next lowest, but to the nearest, number, which is 4. Fitts and Jones (**27**) found this type of error with reference to altimeter reading to be the most common of all instrument reading errors reported by pilots. Although specific data on the use of Navy twin dial instruments are lacking, it would appear that the common design with two dials which rotate in opposite directions past a common index mark would be subject to various reversal errors as well as to the errors just described.

Following the Fitts-Jones study, Grether (**31, 32**) conducted a pencil-and-paper study of alternative ways of representing the broken scale altitude data. The multiple-pointer method now in use was very poor, but a combination of a single pointer for the fine scale running to 1,000 feet and a counter for the coarse record in thousands of feet was about four times as fast and much less subject to 1,000-feet errors. Grether's data established by general opinion that pointers should not be so long that they obscure the graduation marks (**93**). Thus, common practice is to make the pointer just long enough for its tip to touch the inner edge of

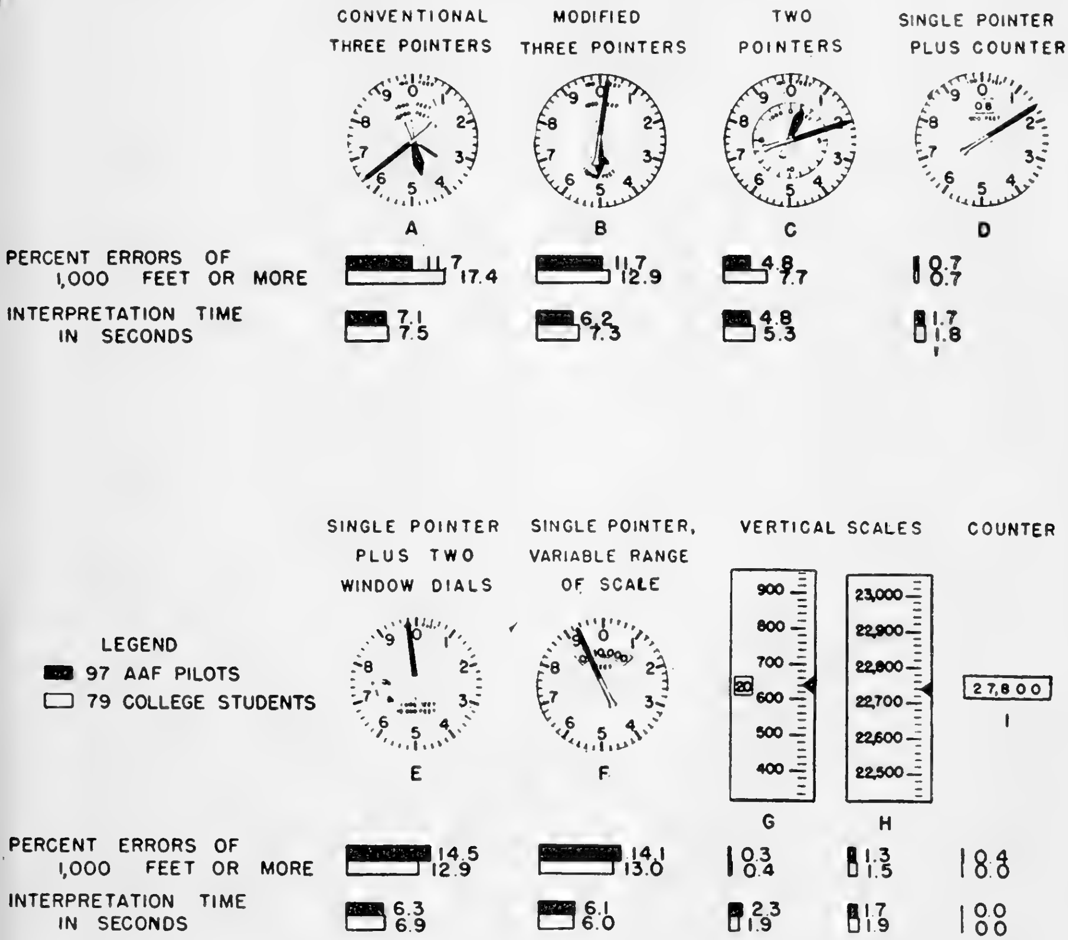

Fig. 11. Reading altitude from different types of instruments. Data from Grether (**31**) for 176 subjects, each of whom made 12 readings on each type of instrument. The results given below each illustration show the percentage of readings which were in error by 1000 feet or more and the instrument interpretation time in seconds. The latter was assumed to be zero for the counter, so counter-test time, essentially number-writing time, was subtracted from all other instrument test times to obtain their interpretation times.

are summarized in Fig. 11. Development along these lines is promising.

Pointer or Index Design. In comparison with problems of scale details, problems related to pointer or index design have received very little research attention indeed. On the subject of pointer length, it has been the shortest scale graduation mark. Regarding pointer width, Loucks (**52**) and Kappauf, Smith, and Bray (**40**) offer evidence which suggests that narrow pointers may be used with no loss in reading accuracy and with the specific advantage that they do not conceal instrument numbering. Loucks also investigated pointer form very briefly, obtaining the result that full-length pointers starting at the center of a circular dial are superior to pointers on which a short length of the tip only is visible. Instruments with twin pointers, as used especially in aircraft to indicate the functioning of two different engines, are difficult and slow to read. The trend at the moment is to replace such dual instruments with pairs of smaller, single-pointer instruments (**59**).

Scale Form. It is the goal of instrument designers to make instruments in such a way that the direction of the indicated change will be easily perceived. The fact that vertical scales seem to be the natural way of expressing changes in magnitude, and of altitude or depth, in particular, has caused this type of scale to receive repeated consideration. Its principal disadvantage is that it is often difficult to mechanize, but it is conceivable that this could be more than outweighed by advantages from the operator point of view. Studies of thermometer type and other vertical scales are currently under way in Great Britain (**68**) and two were recently reported in this country. One of the latter, by Sleight (**70**), unfortunately offers little worthy evidence, since the results were distorted by the fact that the subjects had no opportunity to refixate during the 0.12 second exposure periods which were used. They were of necessity less accurate in reading any scale which was spread out over an extended area. More satisfactory are the data of Grether (**31**) which have already been referred to in Fig. 11. Grether used pencil-and-paper test materials and had his subjects work as rapidly as they could. In this situation, as will be seen from the figure, the speed and accuracy of reading the vertical scales were about the same as for reading the combined circular scale and counter.

If the chosen scale form for any particular application is a circular dial, Christensen's results (**17**) on the errors in reading clockwise and counter-clockwise scales are worthy of note. Christensen used a serial dial-reading situation in which the subject was shifted randomly between clockwise and counter-clockwise dials. Under these conditions, reversal errors, reading the scale in the wrong direction from a numbered graduation, were very frequent. They occurred in about 4% of all the readings made. It is not surprising that a significantly greater number of these reversal errors occurred on the counter-clockwise dial. But by showing that these errors were two and one-half times as frequent as on clockwise scales, the study adds a quantitative emphasis to the warning against using counter-clockwise scales wherever possible.

Some Summary Data on the Accuracy of Making Quantitative Readings

The paragraphs above have dealt with variables which influence, or have been thought to influence, operator performance in taking quantitative data from instruments or indicators. For the most part, the data presented have been stated in relative terms, i.e., design A is better than design B. Some absolute data on reading accuracy are offered now in Table II. This table serves at least two functions. It provides a partial summation of the problems discussed to this point, including representative data from those experiments which have dealt the most systematically with quantitative reading problems. It also provides performance records which should be especially useful to persons who are planning future experiments in this area and who wish to know what levels of reading accuracy they might expect under particular test conditions.

Instruments for Check Reading

Check-reading operations have attracted special attention because of their particular importance for operators working with banks of instruments. Checking such indicators has often been an instrument-by-instrument affair. As a result of the growing number of instruments used on some panels, various schemes have been suggested for simplifying the task of checking them. Such layouts might employ circular scale instruments, each with its scale starting point so chosen that the null or desired indication on every instrument is in the same position—say at the 12 o'clock spot. They might also employ thermometer-type, vertical indicators all nulling at the same height, or a simple set of pilot light, "all or none" indicators.

A test of the relative merits of different ways of arranging dial type instruments for check reading was recently reported by Warrick and Grether (**82**). They used an instrument panel containing sixteen dials 1.75 inches in diameter, arranged in a 4 x 4 square pattern. When the panel was exposed, the subject, responding by voice and switch operation, indicated whether all the dials were set in a prescribed manner or whether one or more of them was off. If the latter was the case, the subjects were required in some tests to report the direction of deviation from the desired setting, i.e., too much or too little. It was not difficult for Warrick and Grether to show that a plan of aligning the null positions for all instruments is superior to having the check positions differently located on different instru-

TABLE II

SOME TYPICAL RESULTS ON THE ACCURACY OF SCALE READING UNDER DIFFERENT EXPERIMENTAL CONDITIONS

Reading conditions	Reference	Subjects	Value of smallest division	Dial diameter	Division mark separation	Readings made to nearest	Allowable error or tolerance	Errors exceeding tolerance
Tachistoscopic exposures, 0.75 second, daylight illumination levels	Loucks (**51**)	60 aircrew cadets	100 rpm.	2.9″	0.26″	0.1 div.	±0.2 div.	24%
	Loucks (**53**)	20 inexperienced aircrew trainees	1 in. Hg.	2.9″	0.13″	0.5 div.	±1 div.	16–20%
			1 in. Hg.	2.9″	0.16″	0.5 div.	±1 div.	21%
			1 in. Hg.	2.9″	0.20″	0.5 div.	±1 div.	12%
			1 in. Hg.	2.9″	0.32″	0.5 div.	±1 div.	9%
Speed and accuracy instructions, daylight illumination levels, dials read in individual presentation	Chapanis (**14**)	5 radarmen and civilians	2°	4.25″	0.072″	0.5 div.	0	8%
			2°	6.25″	0.11″	0.5 div.	0	10%
			1°	10.0″	0.087″	1.0 div.	0	8%
	Grether and Williams (**36**)	80 college students	10	1.9″ to 4″	0.4″ to 1.3″	0.1 div.	0	43%
							±0.1 div.	6%
							±0.2 div.	3%
Accuracy instructions, daylight illumination levels, dials read in groups of 12	Kappauf and Smith (**41**)	6 graduate students	10	1.4″ and 2.8″	0.4″ to 0.9″	0.1 div.	0	14%
							±0.1 div.	1%
							±0.2 div.	1%

ments. Using either the 9, 12, or 3 o'clock position for such alignment was very satisfactory. Of these alternatives, however, the 3 o'clock position was the poorest from the point of view of the accuracy with which operators responded to the direction of instrument deviation when all were not in check. In this respect the 9 o'clock position is thought to have special advantage for check reference because it is "natural" to interpret a high pointer position as meaning too much and a low pointer position as meaning too little of the indicated function.

Perhaps the best way to summarize the advantages of systematic instrument alignment is in terms of some of the specific data reported by Warrick and Grether. If all an operator has to do is indicate whether the instruments do or do not check, he can make this reaction to a bank of 16 aligned instruments in about 0.75 seconds on the average. When the instruments have differently oriented check points, it takes him about 1.6 seconds for the same job. Similarly, if the operator has to respond to the direction of a deviation, not just to the fact that a deviation exists, he saves about two seconds and makes about 50% fewer errors if the check points are aligned than if they are not.

To answer the question whether scale form is important for the speed and accuracy of check reading an isolated or single instrument, Grether and Connell (**35**) undertook a reaction-time test to evaluate five forms of data presentation. Subjects viewed an instrument in repeated exposures and were required to respond to changes from previous indicated values. The instrument forms used were a circular scale with moving pointer, a circular scale moving behind a fixed pointer, two vertical scales, one with a moving and one with a fixed pointer, and finally, a direct reading counter. Unfortunately, the report of this experiment leaves unclear the extent to which the results were influenced by the particular scale values chosen as check points. The circular scale with the moving pointer appeared to be clearly superior to the circular scale which moved behind the fixed pointer, but the significance of the differences between either of these instrument forms and the others was not reported. Until further test data become available it seems appropriate to observe only that the observed differences in check reading performance between the moving pointer circular instrument, the vertical scales and the counter were all quite small.

Problems in the Design of Specific Instruments

Typical of the research which may be needed in order to discover satisfactory designs for instruments which serve specific purposes, is the work which has been done on the class of orientation instruments and on the design of a 2400 military clock. These studies will be reviewed briefly.

Interest in orientation instruments (**30, 34, 38**) stems particularly from the ease with which operators often make reversed interpretations of an indicated change in the orientation of their vessel or aircraft. Such reversed interpretations are very common with the traditional form of aircraft gyro horizon (**11, 12, 54**) on which the horizon bar is stabilized and remains parallel to the true horizon while the airplane figure on the instrument remains in fixed position in the airplane. The pilot, seeing the bar tilt with reference to all other objects in his cockpit, is inclined to interpret that tilt as the angle of his wings—which in fact tilt in the other direction. Out of the study of this and similar orientation devices, Fitts and Jones (**27**) advance the hypothesis that in general the "aircraft reference principle" rather than the "external reference principle" should be followed in the design of instruments for the control of aircraft in situations where split-second reactions are demanded. By the aircraft reference principle they mean the principle of having the moving element of an indicator move in the same direction in relation to the pilot as the aircraft is moving in relation to the ground. Reaction-time data collected by Peterson (**62**) for different aircraft compass designs also support the aircraft reference principle. Although Fitts and Jones anticipate that this principle may not apply to the design of all orientation instruments, they look for research to demonstrate where these exceptions lie.

The design of a clock face to be used in reading time on a 2400 military scale was investigated by Grether (**29**). By using pencil-and-paper test materials he was able to explore the relative merits of eleven different clock designs in reasonably short order. These designs are shown in Fig. 12. Perhaps the most revealing result of this experiment was the wide range of error frequencies found for the different designs. There were only 4% incorrect readings on the best clock, design J, but more than 20% incorrect readings on the poorest designs, C and I. Average time per reading, including time for the subject to write his answer, did not vary so much—from less than 6 to about 7 seconds for the different dials. Specific comparisons within the study showed that it was desirable (1) to graduate every minute, not just every five, (2) to number the minute scale at five-minute intervals, (3) to number all hour positions, (4) to use a 24-hour, single-revolution scale, rather than a 12-hour, two-revolutions scale, (5) to start and end the hour scale at the bottom of the dial, (6) to start and end the minutes scale at the top of the dial in the conventional way. The dial which combined all these features is dial J in the figure.

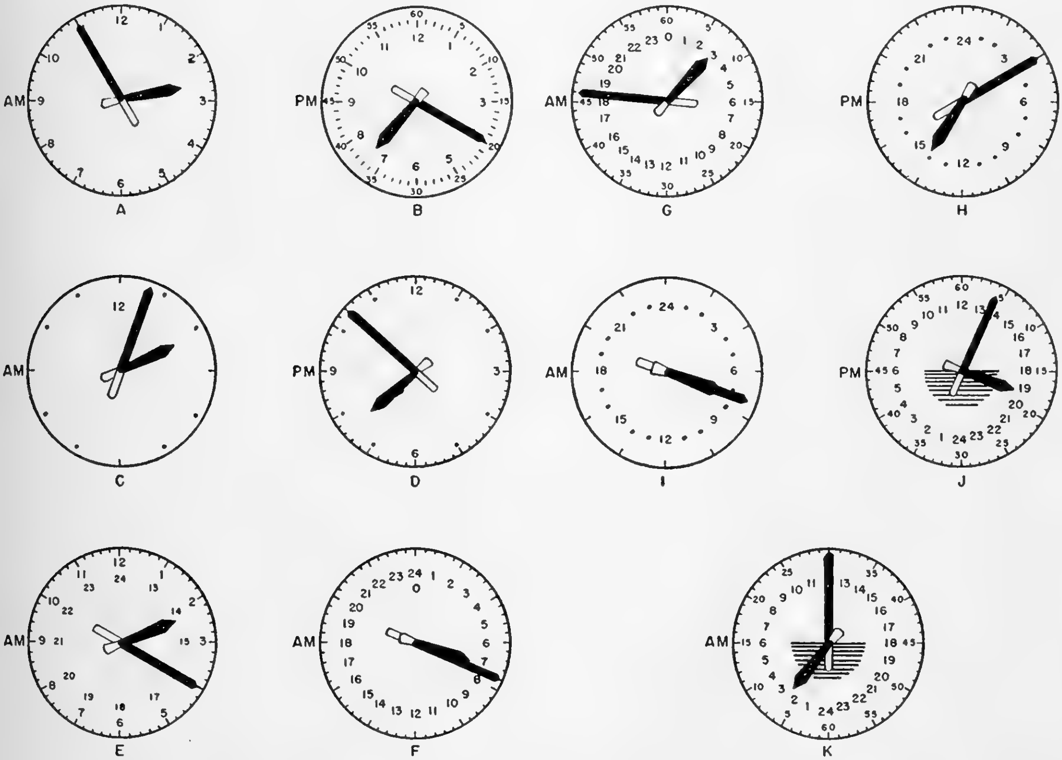

Fig. 12. Experimental designs used in the study of clock dials for reading in the 2400 military time system. From Grether (**29**)

It is clear that problems in the design of other specific instruments will yield to experimental investigation in the same manner as they have for these instruments just described.

Summary

Experiments dealing with the marking, illumination, and detailed design of instruments have been reviewed. It has been observed that in each area of study there are still questions to be answered. Many studies need repetition because of the small number of subjects for whom data have been reported. Other studies have been hastily run and remain incomplete.

Problems of instrument illumination are reasonably well in hand. The merits of counter-type indicators have been well explored. Problems of graduation-mark spacing and interval values are currently under further investigation. Particularly needed are new investigations of direction of contrast effects, of pointer and index designs, and of the relative merits of different scale forms both for quantitative and for check reading operations. Such research, which to date has only been carried out at high levels of illumination, should be extended to determine whether different design requirements prevail for instrument use under low level red illumination.

References

1. Bäckström, H. Über die Dezimalgleichung beim Ablesen von Skalen. *Z. InstrumKde.*, 1930, **50**, 561–575.
2. Bäckström, H. Die Dezimalgleichung beim Ablesen von systemtrischen Skalen, ihre zeitliche Beständigkeit und ihre Abhängigkeit von der Ablesungsgeschwindigkeit, der Übung, u. a. *Z. InstrumKde.*, 1940, **60**, 231–237, 261–271.
3. Banister, H., Hartridge, H., & Lythgoe, R. J. The influence of illumination on visual acuity. *Proc. Opt. Conv.*, 1926, 551–561.
4. Bartlett, F. C. *Instrument controls and display.* GB MRC-RNPRC, Psychol. Lab., Cambridge, RNP 43/81, December 10, 1943, pp. 7.
5. Bartlett, F. C. *Instrument controls and display: Efficient human manipulation.* GB Air Ministry, FPRC 565, December 1943, pp. 6.
6. Berger, C. The dependency of visual acuity on illumination and its relation to the size and function of the retinal units. *Amer. J. Psychol.*, 1941, **54**, 336–352.
7. Berger, C. Stroke, width, form and horizontal spacing as determinants of the threshold of recognition: I. *J. appl. Psychol.*, 1944, **28**, 208–231.
8. Berger, C. Stroke, width, form and horizontal spacing as determinants of the threshold of recognition: II. *J. appl. Psychol.*, 1944, **28**, 336–346.
9. Bray, C. W. *Psychology and military proficiency: A history of the Applied Psychology Panel of the National Defense Research Committee.* Princeton, N. J.: Princeton Univ. Press, 1948, pp. xviii–242.
10. Bromer, J. A. *An experimental compariosn of ultraviolet and indirect red systems of illumination for airplane instrument panels.* USN, NAES, NAMC, Philadelphia, Report TED No. NAM 4422, July 11, 1945, pp. 18, pl. 3.
11. Browne, R. C. *Comparative trial of two attitude indicators. Preliminary note.* GB Air Ministry, FPRC 611, February 1945, pp. 5.
12. Browne, R. C. *Comparative trial of two attitude indicators.* G.B. Air Ministry, FPRC 611(a), April 1945, pp. 5.
13. Byram, G. M. Physical and photochemical basis of visual resolving power. Part II. Visual acuity and the photochemistry of the retina. *J. opt. Soc. Amer.*, 1944, **34**, 718–738.
14. Chapanis, A. *Speed of reading target information from a direct-reading counter-type indicator vs. conventional radar bearing-and-range dials.* Systems Research Field Laboratory, Johns Hopkins Univ., November 1, 1946, pp. 12. (RESTRICTED)
15. Chapanis, A. *The relative efficiency of a bearing counter and bearing dial for use with PPI presentations.* Systems Research, Johns Hopkins University, August 1, 1947. pp. 15–vi. (RESTRICTED)
16. Christensen, J. M. *Psychological factors involved in the design of air navigation plotters.* US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-1d, October 17, 1946, pp. 26.
17. Christensen, J. M. *The effect of the staircase scale on dial reading accuracy.* US AF AMC, Engng. Div., Aero Med. Lab., MCREXD-694-1-P, October 1, 1948, pp. 17.
18. Coover, J. E. Experiments in psychical research. *Psychical Res. Monogr.*, Stanford Univ. Publ., 1917, **1**, 249–265.

19. Craik, K. J. W. *Instrument lighting for night use.* GB Air Ministry, FPRC 342, July 1941, pp. 7; also Psychol. Lab., Cambridge, July 29, 1941, pp. 9.

20. Crouch, C. L. The relation between illumination and vision. *Illum. Engng.* N. Y., 1945, **40,** 747–784.

21. Dunlap, K. *Report of the Highway Research Board.* National Research Council, Division Office, 1932, App. E., p. 3, article 4. (Not seen.)

22. Ferree, C. E., & Rand, G. Intensity of light and speed of vision. II. Comparative effects for dark objects on light backgrounds and light objects on dark backgrounds. *J. exp. Psychol.*, 1930, **13,** 388–422.

23. Ferree, C. E., & Rand, G. Size of object, visibility and vision. *Trans. Illum. Engng. Soc. N. Y.*, 1931, **26,** 820–856.

24. Ferree, C. E., & Rand, G. Visibility of objects as affected by color and composition of light. Part I: With lights of equal luminosity or brightness. *Person. J.*, 1931, **9,** 475–492.

25. Ferree, C. E., & Rand, G. Visibility of objects as affected by color and composition of light. Part II: With lights equalized in both brightness and saturation. *Person. J.*, 1931, **10,** 108–124.

26. Fishenden, R. B. Types, paper and printing in relation to eyestrain. *Brit. J. Opthal.*, 1946, **30,** 20–26.

27. Fitts, P. M., & Jones, R. E. *Psychological aspects of instrument display.* I: *Analysis of 270 "pilot-error" experiences in reading and interpreting aircraft instruments.* US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-12A, October 1, 1947, pp. 47.

28. Fry, G. A., & P. W. Cobb. Visual discrimination of two parallel bright bars in a dark field. *Amer. J. Psychol.*, 1937, **49,** 76–81.

29. Grether, W. F. *Design of clock dials for greatest speed and accuracy of reading in military (2400) time system.* US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-8, October 21, 1946, pp. 16.

30. Grether, W. F. *Discussion of pictorial versus symbolic aircraft instrument displays.* US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-8B, August 4, 1947, pp. 19.

31. Grether, W. F. *The effect of variations in indicator design upon speed and accuracy of altitude readings.* US AAF AMC, Engng. Div., Aero Med. Lab., September 2, 1947, pp. 19.

32. Grether, W. F. *Analysis of types of errors in reading of the conventional three-pointer altimeter.* US AAF AMC, Engng. Div., Aero Med. Lab., March 16, 1948, pp. 23.

33. Grether, W. F. *Design of instrument dials for ease of reading.* SAE Preprint April 1948, pp. 11.

34. Grether, W. F. A novel approach to instrumentation. *Plane Facts*, April 1948 (5-4); pp. 7–9, 28–29.

35. Grether, W. F., & Connell, S. C. *Psychological factors in check reading single instruments.* US AF AMC, Engng. Div., Aero Med. Lab., MCREXD-694-17A, September 20, 1948, pp. 21.

36. Grether, W. F., & Williams, A. C. *Speed and accuracy of dial reading as a function of dial diameter and spacing of scale divisions.* US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-1E, March 31, 1947, pp. 22.

37. Hartline, H. K. *Instrument lighting in aircraft and military vehicles: Its effect on night vision of personnel.* US NRC, CAM Report No. 83, November 1942, pp. 9.

38. Hermans, T. G., & Loucks, R. B. *Annotated bibliography on the psychological aspects of orientation as they relate to aviation.* US AAF AMC, Engng. Div., TSEAA-694-16A, December 1, 1947, pp. 256.

39. Holmes, G. The relative legibility of black print and white print. *J. appl. Psychol.*, 1931, **15,** 248–251.

40. Kappauf, W. E., Smith, W. M., & Bray, C. W. *Design of instrument dials for maximum legibility:* I. *Development of methodology and some preliminary results.* US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-1L, October 20, 1947, pp. 42.

41. Kappauf, W. E., & Smith, W. M. *Design of instrument dials for maximum legibility:* II. *A preliminary experiment on dial size and graduation.* US AAF AMC, Engng. Div., Aero Med. Lab., MCREXD-694-1-N, July 12, 1948, pp. 16.

42. Kappauf, W. E., & Smith, W. M. *Design of instrument dials for maximum legibility:* III. *An evaluation of reading difficulty for different parts of a dial.* US AAF AMC Engng. Div., Aero Med. Lab., 1949, in preparation.

43. Kaufman, E. L., Reese, T. W., Volkmann, J., & Rogers, S. *Accuracy, variability and speed of adjusting an indicator to a required bearing.* Systems Research, Johns Hopkins Univ., September 2, 1947, pp. 37.

44. Kendall, M. G., & Babington Smith, B. Second paper on random sampling numbers. *J. Roy. Stat. Soc.* Suppl., 1939, **6,** 51–61.

45. Knauft, E. B. *The design, attachment and precision of an auxiliary range dial for the Mark 4 Radar.* US NDRC, Applied Psychology Panel, OSRD Report No. 3906, July 17, 1944, pp. 8, Restricted.
46. Lauer, A. R. Factors which influence visibility in daylight and under artificial illumination. *Proc. Iowa Acad. Sci.*, 1933, **40,** 185.
47. Lawrence, M., & MacMillan, J. W. *Annotated bibliography on human factors in engineering design.* USN, BuMed., Aviat. Br. Res. Div., NavMed 1079, Project X-651(AV-340-a), February 1946, pp. ix–209. (RESTRICTED)
48. Lawrence, M., & Peckham, R. H. *Evaluation of directional gyro control unit GE Model SKC 3AC.* Letter to Cmdr. O. W. Chenault (MC) USN, BUMED-X-RCO Ia, July 19, 1945.
49. Lee, R. H., & Finch, E. M. *Instrument lighting and low level illumination in submarine conning towers.* USN, NMRI, Research Project X-380, Report No. 1, January 17, 1945, pp. 16. (RESTRICTED)
50. Loucks, R. B. *Legibility of aircraft instrument dials: The relative legibility of tachometer dials.* US AAF, School of Aviation Medicine, Randolph Field, Texas, Project No. 265, Report No. 1, May 30, 1944, pp. 8, pl. 3.
51. Loucks, R. B. *Legibility of aircraft instrument dials: A further investigation of the relative legibility of tachometer dials.* US AAF, School of Aviation Medicine, Randolph Field, Texas, Project No. 265, Report No. 2, October 27, 1944, pp. 4, pl. 2.
52. Loucks, R. B. *Legibility of aircraft instrument dials: The relative legibility of various climb indicator dials and pointers.* US AAF, School of Aviation Medicine, Randolph Field, Texas, Project No. 286, Report No. 1, November 25, 1944, pp. 5, pl. 5.
53. Loucks, R. B. *Legibility of aircraft instrument dials: The relative legibility of manifold pressure indicator dials.* US AAF, School of Aviation Medicine, Randolph Field, Texas, Project No. 325, Report No. 1, December 7, 1944, pp. 4, pl. 4.
54. Loucks, R. B. *Evaluation of aircraft attitude indicators on the basis of Link Instrument Ground Trainer performance.* US AAF, School of Aviation Medicine, Randolph Field, Texas, Project No. 341, Report No. 1, June 22, 1945, pp. 15, pl. 2.
55. Luckiesh, M., & Taylor, A. H. Visual acuity at low brightness levels. *Amer. J. Ophthal.*, 1944, **27,** 53–57.
56. McFarland, R. A. *Human factors in air transport design.* New York: McGraw Hill, 1946. Pp. xix–670.
57. McGehee, W. *Comparative study of pilot fatigue resulting from extended instrument flight using the standard AAF and British instrument panels.* USN, Bu-Med., Final report, Project TED No. ATL-R-601, not dated, pp. 48.
58. Maier, E. Zur Bestgestalttung der zifferblätter von stopphuhren. *Industr. Psychotech.*, 1931, **8,** 97–113.
59. Moore, R. C., & Tate, G. F. *Engine instrument simplification program: "Single" vs. "dual" instruments.* US AAF AMC, Engng. Div., TSEPE-655-1478, November 29, 1946, pp. 47.
60. Paterson, C. C., Walsh, J. W. T., & Higgins, W. F. An investigation of radium luminous compound. *Proc. phys. Soc.*, London, 1917, **29,** 215–249.
61. Paterson, D. G., & Tinker, M. A. *How to make type readable.* New York: Harper and Bros., 1940. Pp. xix–209.
62. Peterson, O. Über die Sinnfälligkeit von Blindflügesgeräten. *Industr. Psychotech.*, 1939, **16,** 225–239.
63. Reed, J. D., & Bartlett, N. R. *The accuracy of range estimation on a PPI with respect to the distance of the target from the range rings.* Systems Research, Johns Hopkins Univ., November 1, 1947, pp. 9. (RESTRICTED)
64. Reese, T. W., Volkmann, J., Rogers, S., & Kaufman, E. L. *Special problems in the estimation of bearing.* Systems Research, Johns Hopkins Univ., January 15, 1948, pp. 46.
65. Richtmyer, F. K., & Howes, F. L. A method of studying the behavior of the eye under different conditions of illumination. *Trans. Illum. Engng. Soc.*, 1916, **11,** 100–113.
66. Riggs, L. A. *The development of a counter or odometer-type scale to replace the micrometer scale of the panoramic telescope M-12.* US NDRC, Applied Psychology Panel, OSRD Report No. 4390, November 28, 1944, pp. 10. (RESTRICTED)
67. Rogers, S., Volkmann, J., Reese, T. W., & Kaufman, E. L. *Accuracy and variability of direct estimates of bearing from large display screens.* Systems Research, Johns Hopkins Univ., May 15, 1947, pp. 41.
68. Ruffell Smith, H. P. *The design and display of pilots' instruments.* GB Air Ministry, Flying Personnel Research Committee, FPRC 692, March 1948, pp. 19. (RESTRICTED)

69. Schulz, H. Die Ermüdung des Auges. *Industr. Psychotech.*, 1925, **2**, 5–12.
70. Sleight, R. B. The effect of instrument dial shape on legibility. *J. appl. Psychol.*, 1948, **32**, 170–188.
71. Smith, W. M., & Kappauf, W. E. *Studies pertaining to the design and use of visual displays for aircraft instruments, computers, maps, charts, and tables: A bibliography.* US AAF AMC, Engng. Div., Aero Med. Lab., TSEAA-694-1G, June 30, 1947, pp. 25.
72. Spragg, S. D. S., & Rock, M. L. *Dial reading performance as related to illumination variables.* I. *Intensity.* US AF AMC, Engng. Div., Aero Med. Lab., MCREXD-694-21, pp. 32.
73. Spragg, S. D. S., & Rock, M. L. *Dial reading performance as related to illumination variables.* II. *Spectral distribution.* US AF AMC, Engng. Div., Aero Med. Lab., MCREXD-694-21A, pp. 26.
74. Taylor, C. D. The relative legibility of black and white print. *J. educ. Psychol.*, 1934, **25**, 561–578.
75. Tinker, M. A. The relative legibility of the letters, the digits, and of certain mathematical signs. *J. gen. Psychol.*, 1928, **1**, 472–496.
76. Tinker, M. A. Relative legibility of modern and old style numerals. *J. exp. Psychol.*, 1930, **13**, 453–461.
77. Tinker, M. A. The effect of illumination intensities on speed of perception and upon fatigue in reading. *J. educ. Psychol.*, 1939, **30**, 561–571.
78. Troland, L. T. Analysis of the literature concerning the dependency of visual functions upon illumination intensity. *Trans. Illum. Engng. Soc.*, 1931, **26**, 107–196.
79. Vernon, M. D. *Scale and dial reading.* Unit in Applied Psychology, the Psychological Laboratory, Cambridge, A.P.U. 49, June 1946.
80. Verplanck, W. S. *A brief experiment on the legibility of the TDC under several conditions of illumination and adaptation.* U.S.N., U.S. Submarine Base, New London, Conn., Medical Research Department, (AV-273-p). Final report on BuMed Project X-519, March 15, 1946, pp. 8.
81. Walls, G. S. Factors in human visual resolution. *J. opt. Soc. Amer.*, 1943, **33**, 487–505.
82. Warrick, M. J., & Grether, W. F. *The effect of pointer alignment on check reading of engine instrument panels.* US AAF AMC, Engng. Div., Aero Med. Lab., MCREXD-694-17, June 4, 1948, pp. 33.
83. Weston, H. C. *The relation between illumination and industrial efficiency.* I. *The effect of size of work.* Medical Research Council Report. London: His Majesty's Stationery Office, 1935.
84. Weston, H. C. *The relation between illumination and visual efficiency—the effect of brightness contrast.* GB MRC, Industrial Health Research Board, Report No. 87, H.M. Stationery Office, London, 1945.
85. Wilcox, W. W. The basis of the dependence of visual acuity on illumination. *Proc. nat. Acad. Sci.*, Wash., 1932, **18**, 47–56.
86. Wilcox, W. W. An interpretation of the relationship between visual acuity and light intensity. *J. gen. Psychol.*, 1936, **15**, 405–435.
87. Yule, G. U. On reading a scale. *J. Roy. Stat. Soc.*, 1927, **90**, 570.
88. *Final report on studies of errors in field artillery practice.* U. S. Army Service Forces, Armored Medical Research Laboratory. Project AMRL No. 37, June 13, 1945, pp. 5.
89. *Instrument lighting in tanks.* G. B. Medical Research Council, Military Personnel Research Committee. Physiological Laboratory, AFV School, Lulworth, PL 19, November 2, 1941, pp. 2.
90. *Minutes of the 6th meeting of FPRC.* G.B. Air Ministry, Flying Personnel Research Committee. FPRC 63, October 4, 1939, pp. 14. (RESTRICTED)
91. *Numerals and letters—standard form of aircraft instrument dial.* U.S. Army-Navy. Army-Navy aeronautical design standard, AND10400, October 11, 1944, p. 1.
92. *Report by subcommittee of the Army-Navy NRC Vision Committee on standards to be employed in research on visual displays.* U.S. Army-Navy NRC Vision Committee. Minutes and proceedings of the 19th meeting. May 27–28, 1947, pp. 13–18. (RESTRICTED)
93. *Report of Army-Navy Conference No. 13–43 on Army-Navy Aeronautical Standardization of requirements for Simplification of Instrument Dials.* U.S. Navy, Naval Aircraft Factory, Philadelphia, Pa., March 4, 1943, pp. 10.
94. *Report on type for government printing.* His Majesty's Stationery Office, 1922.

Chapter 4

VISIBILITY ON RADAR SCOPES

STANLEY B. WILLIAMS[1]

The Johns Hopkins University

Introduction

A novice taking his first look at a radar scope finds that it yields no meaningful information. Even after considerable practice and training, radar operators find that scope reading is not easy. Part of the difficulty lies in the fact that cathode-ray tubes are at best rather poor purveyors of visual information, being made, as they are, of a combination of curved glass surfaces, uncertain chemicals, and delicate electrical circuits. Other difficulties arise if the user does not properly adjust certain aspects of scope function and presentation. The possibilities of improvement are great, and many of them are contingent upon knowledge of the human operator and of his capacity for visual discrimination.

Cathode-Ray Tubes Used as Radar Scopes

General Design

A cathode-ray tube (CRT) is a type of vacuum tube designed for the visual display of electrical events. The display is a *phosphor screen*, a thin coating of inorganic salts on the inside surface of the top of the glass tube. From an external source a high potential is put on the anode to attract the electrons being emitted by the cathode to the region of the phosphor screen. All except the screen is called the *electron gun*. The number of electrons it fires toward the screen is measured as beam current and is primarily governed by variations in bias applied to the control grid of the tube. The direction of the beam is governed by either magnetic or electrostatic deflecting coils collared around the neck of the tube, one set of coils being used for vertical deflection, the other for horizontal, the two together being sufficient for locating the luminous spot at any position on the phosphor screen.

Types of Cathode-Ray Tubes

There are two basic types of cathode-ray tubes: *deflection modulated* and *intensity modulated*.

In the deflection-modulated type, the information or "signal"[2] to be displayed is put into the CRT gun as a spatial deflection of the electron beam. This is the type used in the familiar laboratory oscilloscope, and in radar it is called an *A-scope* (or A-scan). The observer sees the signal as a momentary vertical line or "spike," whose extent is usually a linear function of signal strength.

In the intensity-modulated type, the signal is put in as a variation of the beam current, i.e., the beam is momentarily intensified so that more electrons reach the screen and locally excite a spot to a brightness greater than its surround. Beam intensification is a linear function of the voltage drop in the grid bias caused by the signal. The observer sees the signal as a local bright spot whose area is determined by other fixed parameters and whose incremental brightness *may* not be—and usually is not—a linear function of signal strength, depending on the type of phosphor,

[1] The writer is deeply indebted to Beverly Richards for invaluable bibliographic and research assistance.

[2] The following words are used almost interchangeably to denote the actual or potential screen appearance of a radar echo: echo, trace, image, target, signal, blip, and pip.

its excitation history, the amount of space-charge collected near the screen capable of repelling electrons, etc. Examples of this type in radar are the *PPI*, or Plan Position Indicator, the *B-scope* and *Eagle scans*. Intensity modulated scopes usually use medium or long-persistent phosphors, i.e., phosphors which glow after electronic bombardment has ceased.

Types of Radar Displays

There are three useful dimensions on a scope: the two space coordinates and intensity. Displays are denoted by type according to the use made of these dimensions.

A-Scan or A-Scope

This uses only the space coordinates, by deflection modulation, and keeps intensity essentially constant.[3] The horizontal dimension is a time sweep; since radar target distances are measured only by the time consumed by a pulse's travel to the target and back, they can be displayed as distances along a calibrated baseline. Usually this is made to read from left to right, the left side of the scope representing the pulse source and the right side the limit of the range being scanned. The seeing task places a heavy burden on acuity functions, on the perception of form, on judgments of small differences in linear extent, and on linear interpolation.

B-Scan

The radar B-scope uses all three dimensions. The vertical dimension measures sweep time or range (distance from source). The horizontal dimension represents the successive angular orientations of the rotating antenna, and therefore indicates the relative bearing of the target. The intensity or contrast of the pip is a rough index of signal strength. The seeing task usually is an alignment of the pip with a reference spot or line, but brightness fluctuations must also be appreciated.

[3] Pictorial descriptions of the several types of radar scan are given in Chapter C.

PPI-Scan

The Plan Position Indicator presents a circular (polar coordinate) map with the center the position of the radar antenna (Fig. 1). Radial distance is range; angular orientation is bearing; and pip intensity is signal strength. The angular width of the pip may also furnish useful information about the echoing object, but it mainly reflects only the dispersion (lobe) pattern of the transmitted pulse energy. As the antenna rotates, the sweep trace follows it, painting a succession of lines whose discreteness varies with (a) the frequency of the pulse transmission (Pulse Repetition Frequency, or PRF) and with (b) the antenna's rotation rate. Usually the combination of the two is sufficient to paint in a fairly uniform background, but one which always retains a brightness gradient, highest at the sweep line and fading off around the circle behind it—much like a circular optical wedge. The critical seeing task is brightness discrimination, especially for weak signals, but form discrimination and acuity increase in importance with the amount of "noise" (clutter of unwanted pips or pseudo-pips due to atmospheric interference and intra-tube emission).

Other Scans

Most of the others are classifiable as Eagle scans, portions of a PPI-scan, for which the seeing task is similar to that of the PPI. (See Chapter C.) Television, of course, is a special case of intensity modulation in which the vertical and horizontal dimensions are made to correspond to the dimensions of the object being televised.

Types of Phosphor Screens

Upon excitation, either by electrons or by light of short wavelengths, the screen material fluoresces momentarily. The luminescence may last only long enough to be seen or it may excite another phosphor component whose glow may be long persistent. Both color and afterglow can be varied during screen manufacture by appropriate selection of the phosphor components. This variation is largely empirical; rationalization of phosphoric luminescence in terms of atomic weights or any other scheme has defied accomplishment. Leverenz of RCA states, for example, "There are, as yet, no theories which give quantitative agreement with experiment. Because it is only rarely possible to predict the luminescent properties of proposed new phosphors, research in this field must be guided by 'scientific intuition' rather than logic" (**29**). In cooking up a batch of phosphor the manufacturer pays attention to crystal size, valence, impurities and other factors; and in coating the glass tube, the thickness of the coating, degree of optical contact, and even thickness of the glass must

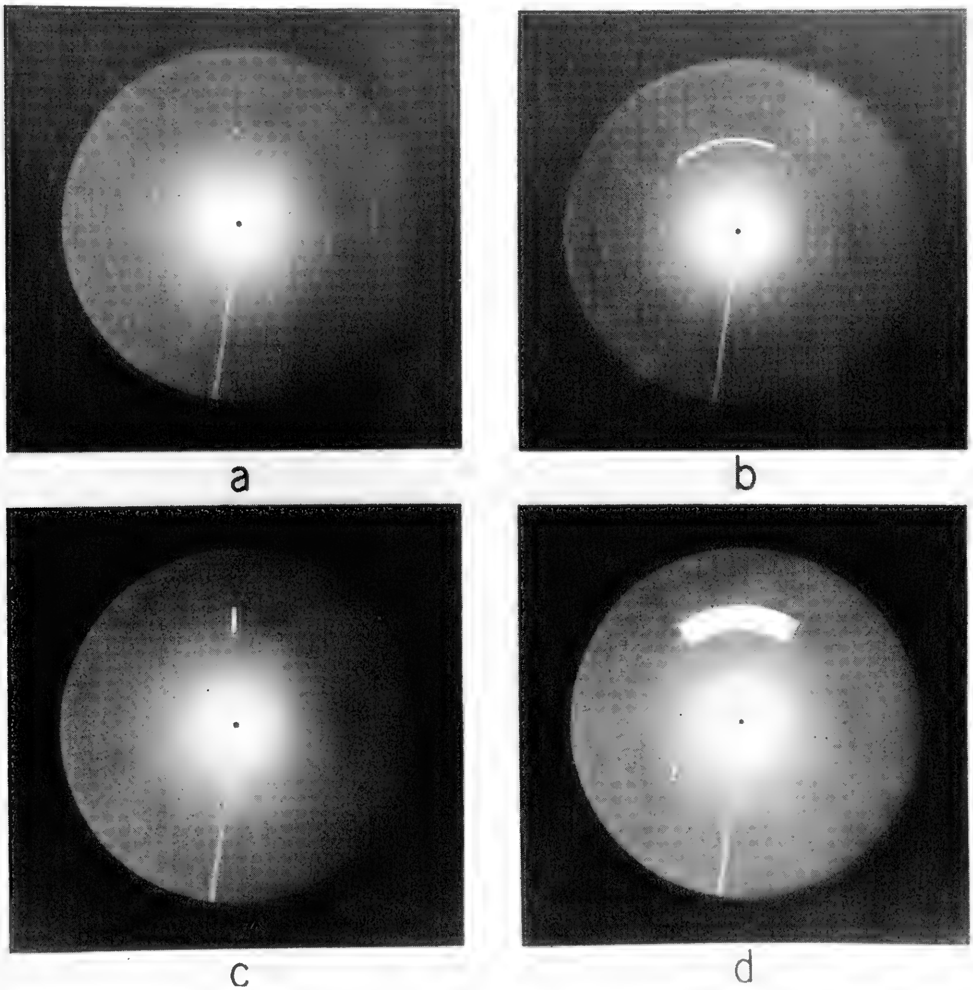

Fig. 1. Photographs of PPI scope and single pip

The photograph shows the PPI scope actually used in several experiments at The Johns Hopkins University. It is a 7BP7 tube, made by RCA; it has an effective diameter of 6 inches. The radial line is the rotary *sweep;* the concentric lines are electronic reference rings used in estimating range. The pip in these pictures is seen at slightly more than 10 miles, counting each ring as 5 miles. Pip sizes shown are (a) 1 μs x 1°, (b) 1 μs x 60°, (c) 30 μs x 1°, and (d) 30 μs x 60°.

be watched. Any of these can affect visibility and wide variations may exist among screens of identical code name. There is, therefore, no sure way of duplicating most screens for exacting research purposes. For this reason, no complete listing of common phosphors will be here attempted. Leverenz (**29**) has a compendium of the properties of 23 high efficiency and 22 low efficiency phosphors which have been studied in the laboratory and, in other articles (**30, 31**), presents the properties of the commonly manufactured screens. Of these, only about five are much used in radar work. They are divided into two major classes with respect to persistence of afterglow.

The short persistence class has one outstanding member:

P1. Used on most laboratory oscilloscopes, in some rapid-scan radar. Usually used in A-scan. Light green; decays to 1 per cent in about 50 milliseconds. Zinc-silicate activated with manganese ($ZnSiO_4$:Mn). Very stable, often used as a reference phosphor for relative energy and efficiency. Narrow emission band with well-defined peak at 523 mμ.

The long persistence class are called "memory" screens; they not only preserve strong echoes for many seconds but help to integrate weak, recurring signals (see Sutro, **46,** and Brill, **8**). Examples are:

P7. The most widely used long persistence phosphor today, especially on PPI's. Consists of two phosphor layers, hence a "cascade" screen. The first layer (ZnS:Ag) fluoresces blue on electronic excitation with a peak emission at about 435 mμ. The bright blue itself decays in a few microseconds, but its light energy excites the second layer (ZnCdS:Cu) to a whitish yellow with a peak emission at about 570 mμ. Very variable. (See Nottingham, **41.**)

P14 Similar to P7 except that second layer (ZnCdS:Ag) glows orange rather than yellow, with peak emission at about 600 mμ (See Innes, **25.**)

P10. Sometimes called a scotophor (KCl) rather than a phosphor, because it darkens on excitation rather than brightens. Extremely long-persistent; requires heat for erasing.

In addition, a medium persistence screen, the P12, may be mentioned. It is a single layer screen (ZnF_2:Mn) which glows orange, has a peak emission at about 589 mμ, and is often used in fire-control radar.

Methods of Research on Visibility

Equipment

Either real or simulated radar scopes have been used. An entire real radar system has only the advantage of realism; usually it is a poor research device.

Real Radar

Naval fleet units have made comparisons between complete radar systems installed in adjacent ships. A clear superiority of one system over another can be established by such means, though the reasons for it often have to be sought in the laboratory. Or, statistical selection of radar contacts can be drawn from operational records and often it is possible to make useful comparisons of systems, operators, or scopes. The method is laborious and expensive.

Artificial Radar

Still retaining visual realism, several groups of scientists, (notably those at the Radiation Laboratory, Massachusetts Institute of Technology; Naval Research Laboratory; Division of Radiophysics of the Australian Council for Scientific and Industrial Research; and Systems Research Laboratory of the Johns Hopkins University) have employed an artificial radar set with simulated signals but real scopes with radar scans. This method is costly, but signal generators, oscillators, and meters cost less than ships. Even so it cannot be undertaken by a small laboratory. In spite of the expense, the method is probably the best compromise between scientific control and realism.

CR tubes but No Simulation of Radar Scan

Another method makes use of CR tubes but makes no attempt to simulate scanning or real targets. The signal is a "raster," a small rectangle of the phosphor screen electronically excited at controllable intervals and intensities. Though little used as yet in visual studies, it was successfully employed by Bartlett (**3**) to study flash frequencies and by Hanes (**18**) to study comparative brightness judgments. More commonly, the viewer is a photocell or a phototube. However, data can be correlated with knowledge of visual requirements The correlation is more often implicit than explicit and is inclined to be based on estimates of a "photopic eye."

Optical Simulation

The work done in England by Hopkinson (**24**) employed only an optical imitation of a radar scope. Such studies require an assumption of successful imitation. Often this can be plausible. Some of the optical imitations developed for training purposes may even be adaptable to research (**33, 36, 47**). Data from imitations are yet to be compared with those obtained from real scopes.

Physical Measurements

The critical stimulus for vision is difficult to measure, especially on intensity modulated scopes. In A-scopes, the deflection is discriminated by its relative height and by the width of the gap in its baseline. A psychophysical study of each cue would require their independent variation, a nearly impossible task on a real scope. The practice has been to measure the voltage or power of the deflection, rather than its extent, relative to some selected "average" noise voltage or power. From a purely visual standpoint the threshold A-scan deflection is some mongrel form of a visual acuity, and the signal-to-noise ratio is meaningless. In intensity-modulated scopes, detection requires a combination of form and brightness discrimination, and the signal-to-noise power is an equally inappropriate unit. For brightness discrimination experiments one would like to use brightness units. Because of variations in electrical circuits and because of the marked effects of afterglow (in the memory screens) on brightness level, it is virtually impossible to control screen brightness precisely by means of meters or scopes that register only volts or amperes. Measurements made at Systems Research Laboratory, for example, showed a difference of about seven volts for the visual cut-off bias between an old tube and a new tube of identical type. Over months of daily use, this bias in the new tube gradually receded in the direction of the old tube. This means that it is never safe to compare tubes or conditions by means of a voltmeter alone. It became necessary for the workers at Systems Research Laboratory to depend on a visual estimate of sweep cut-off as a reference point for determining background brightness (**55**). Even using a visual reference, brightness *differences* must be measured by meters, for subjective estimates of supra-threshold brightness levels are not very accurate. It would be desirable to have a physical instrument sufficiently adaptable to use continuously as a monitor for the brightness of the scope, but such instruments are not available. The closest approximation is an electron multiplier tube, (usually either an RCA 931A or 1P22 multiplier phototube) in conjunction with a dc amplifier. These tubes can amplify weak photo-currents as much as 200,000 times, but they lack two desirable features: their lower limit of light sensitivity is between 0.01 and 0.1 footlambert, which is still about three log units above the eye's threshold; they are likely to be cumbersome to focus and to read, without extremely elaborate and expensive accessories.

In the absence of a brightness scale, thresholds of signals have conventionally been expressed in decibels of attenuation of a reference signal voltage or power, assuming constant noise. The scale is arbitrary with respect to absolute values of energy but is entirely meaningful for comparative purposes. Inasmuch as satisfactory agreement has never been reached on the definition of noise, different investigators select different noise levels as reference points. This leads to discrepant statements as to how far below average noise level a signal can be seen. Since visibility measures will have to be on an arbitrary scale for a long time, it is a pressing problem to reach agreement on the reference intensities.[4]

Psychophysical Methods

Memory screens pose a new problem as to the actual procedure of taking a threshold. Their decay characteristic precludes immediate re-excitation of a screen spot which is still glowing. It usually is necessary to adopt a sequence of dimmer-to-brighter for the experimental runs on a given day although it is possible to deactivate the phosphor by red light between sessions. Traditional methods of psychophysics are applicable with some modifications.

The Method of Average Error, wherein the subject adjusts a signal until it appears "just visible," works out very well, provided observers are trained enough to be capable of maintaining a constant criterion of judgment. The interval of uncertainty, the range over which adjustment is made prior to settlement, is usually no more than a few db, which is sufficiently small to minimize decay and build-up effects. The method has the great advantage that it does not require extended observations over long periods during which electrode voltages are likely to drift.

In a modified Method of Limits, the subject simply watches a scope initially devoid of signals while the experimenter gradually introduces a pip. The subject may or may not know the location of the pip before hand, depending on the nature of the experiment. One obtains an *appearance threshold.* The method is similar to the sighting of a target in operational practice, but apparently was not used in radar research until Garner. A description of its use with multiple targets, staggered in strength and therefore in order of appearance, can be found in Garner and Hamburger (**15**).

A modified version of the Method of Constant Stimuli is usable though tedious. However, because of phosphor decay and build-up, it is impossible to employ a truly random order of signal intensities unless the signal is shifted in scope position from trial to trial. This tends to confound the threshold and obviate much of the method's traditional precision. Using an unconventional version of the method, the M. I. T. recearchers (**1**) discovered that the operator's guesses of marginal stimuli improved in acsuracy as stimulus strength increased—a fact known for nearly a century by psychophysicists—and that one had to be arbitrary in the selection of a probability criterion. Not being acquainted with psychophysics, they decided to use the 90 percent point on a fitted linear function instead of the traditional 50 percent point on a fitted ogive, a difference which happily is negligible for most engineering purposes. The method may also require a circuit stability not easily achievable with electronic equipment.

Kinds of Factors Influencing Visibility

Physical Characteristics of the Glass Tube

The light generated by a spot on a phosphor screen travels many paths; only about 25–35 percent of it is transmitted through

[4] The decibel is one-tenth of a bel, which in turn is a power ratio of 10 to 1. Although customarily applied to sound intensities the decibel can be applied to any kind of energy and is commonly used in electronics. To make sense, a reference intensity must be defined "absolutely," that is, in some other unit. In sound, an arbitrary reference pressure has been agreed on by scientific workers. No such agreement exists among radar investigators and each laboratory chooses its own. No matter what the reference signal voltage might be, however, use of the decibel unit means a logarithmic scale. Ten db is $\frac{1}{2}$ log unit in voltage; one db is a voltage ratio of approximately 1.13 to 1; six db is approximately 2 to 1. For example, to double a signal voltage, we increase it by six db, which is accomplished by *reducing* the *attenuation* of the reference signal by six db.

the glass tube in the direction of the observer. Some is reflected back into the bulb and finally ends up at points on the screen remote from the spot. Law (**28**) lists the following as factors unfavorable to good contrast: (a) bulb wall reflections, (b) screen curvature, and (c) halation, i.e., internal reflections in the glass adjacent to the spot which give rise to concentric rings or "halos" of light whose separation is proportional to the thickness of the glass. He compared these factors with two others, viz., room illumination and stray electrons, and concluded that halation was much the most important factor in CRT design. Stray electrons are no longer bothersome because of improved electron guns. Bulb wall reflections apparently are minimized by their being diffused, a consequence of proper bulb shape and internal coatings. (A comment on this is of interest, because the cone shape of American tubes may or may not be the best possible from the diffusion standpoint alone and one wonders why the German tubes are so nearly cylindrical. One would think that a more nearly spherical bulb would be preferable, provided other design considerations do not outweigh it.) Screen curvature is somewhat more important than bulb wall reflections but is still not a limiting factor in contrast, except insofar as it may enhance halation. This matter apparently has not been studied for its relevance to radar. Halation alone, Law found, could limit contrast to a ratio of six to one; but it can be materially reduced by the introduction of light-absorbing material into either the glass face or the binder of the phosphor. In amounts of 10 to 20 percent, this absorbing action will have a negligible effect on the rays directly transmitted through the glass to the eye, but will largely eliminate the weaker halation reflections. This seems to be a good enough idea when considering television's high-brightness CRT's, as Law was, for these tubes produce more than enough light for good vision anyway, but whether the principle should be applied to low luminosity radar scopes is another matter. The writer has been able to find no evaluation of the question.

An additional factor has been studied by Bachman (**2**), namely, the effect of non-reflecting glass surfaces. Tests made with non-reflecting lead glass yielded contrasts superior to those of good-quality untreated glass. Whether the same would hold for the several other types of non-reflectance treatments is not known. In fact, we have not been able to ascertain whether the commercial CRT faces are actually treated.

The new metal-backed screens may as well be mentioned here, because they represent one attempt to save some of the light ordinarily lost in internal reflections. Bramley (**7**) has confirmed the work of Bachman (**2**) and of Epstein and Pensak (**12**) which showed that a thin metallic coating on the gun side of the phosphor screen actually reflected back toward the observer's side a sizeable fraction of this light, approximately doubling the light output of the tube. The metal backing acts like a mirror, but one, fortunately, which is pervious to electrons. Total light increase is greater than contrast improvement, but both together *ought* to be beneficial for detecting weak radar signals. No known experimental test has been made. In addition, the luminous efficiency is said to be increased, at least at higher anode voltages, because the metal is an aid in reducing the space-charge on the screen which normally prevents faithful following of the anode potential. If true, the metal backing deserves a real try-out on radar CRT's. At this date it is only just beginning to be used in the television tubes.

Phosphor Screens

The color and decay characteristics of commercial screens allow a narrow selection for the researcher. If the psychology or physiology of phosphors is to be known, it may eventually be necessary to carry out direct visual studies on phosphor materials. So far this has not been done, at least in any systematic fashion. The current method of evaluating phosphor materials is by photocell or multiplier phototube, the material being excited by light of short wavelength. Good descriptions of the method are found in Hopkinson (**23**) and in Hardy (**20**). A similar method is used in evaluating completed CRT screens, except that here the excitation is by electrons. Measures are made of spot diameter, spot brightness, decay times, build-up on successive excitations and others which, together, constitute a fairly complete battery of quantitative tests which ought to predict operational performance. Good accounts of the method are given by Johnson and Hardy (**26**), by Garlick, Henderson, and Puleston (**14**), and by Nottingham (**38, 39, 41**). The scan pattern, of course, is not a radar one, but a small test raster. Scientific caution dictates that photocell measures ought properly to be regarded as only a set of pre-tests which may not always predict very accurately the luminances which are critical for effective visual use. That they actually do fail is witnessed by Nottingham (**41**), who concludes that, "It has not been possible to bridge the gap between objective measurements and radar tube applications and it is unlikely that this gap will be closed in the near future" (**41,** p. 15). He was referring specifically to a comparison of the "cyclic excitation" method of the British—a photocell index of phosphor decay—and the "build-up ratio" method of the Americans—an index of cumulative brightness with successive excitations. Neither method measured directly a signal-background contrast of the sort found in operational radar, mainly because they were not effectively simulating noise excitation. Consequently no completely satisfactory test of phosphor performance has been developed to use in manufacturing control. The ultimate test is human vision; but though this is admittedly cumbersome as an acceptance test in a tube plant, certainly it might be possible to construct a visually valid photocell "aptitude" index. Nottingham's problem *may* be soluble by the statistical methods familiar in aptitude testing.

The measurement of luminescence during operational scan is almost too difficult to accomplish, although a partially successful attempt has been made at Systems Research Laboratory by Hamburger and King (**17**) whose records proved extremely useful in the interpretation of human visibility data (**54, 55**).

The properties of an efficient screen are difficult to specify. Screen thickness, for example, goes through a definite optimum, but the optimum in turn depends on amount of binder and method of processing, as well as on the working voltage (**23**). The junction of the phosphor and the glass and the elimination of metallic impurities are extremely important. Some appreciation of how tubes are made is almost a requisite for visual research. As an empirical guide, a population study of manufacturer's tube types, using visual performance measures, is urgently needed.

A great contribution to phosphor design might be made by psychology and physiology. For example, manufacturers have often assumed a constant of $\frac{1}{10}$ second (**30, 31**) as the value for "visual persistence." We know, of course, that no theory of temporal integration of the eye is as yet completely founded; yet, some of the empirical data of the flicker studies might be applicable, and certainly the needs of radar should stimulate new research. In England, Hopkinson (**24**) felt it necessary to carry out experiments not only on flicker but on movement of flashes and their apparent brightness in order to arrive at some rational guide to CRT construction. Recently research has been undertaken by Bartlett (**3**) and Sweet and Bartlett (**48**) which has a direct bearing on the problem. Reliable data on integration time as a function of stimulus size, brightness and movement and of background area, texture and brightness should be brought to bear on special problems. In passing, it may be noted that the development of reliable scales of subjective brightness (**18**) may aid materially in the rationalization of these visual functions.

Basic Parameters of Electrical System

In the whole train of electrical events from the initial broadcast of the pulse of energy, through its echoing and reception and amplification, there are numerous stages at which influence on the final pip can be exerted. Most of these affect its intensity; some may affect its shape. Many of them—particularly those which affect signal strength—will ultimately determine visibility in about the same fashion, no matter what the form of the indicator tube's display. Some of the relationships are critically dependent on display forms. Even leaving out minor factors such as construction imperfections, tube life, and such, there is a long list of system parameters that presumably *could* affect signal detection. The major ones are:

1. Length of pulse (duration)
2. Pulse repetition frequency
3. I-f bandwidth
4. Video bandwidth
5. Sweep rate
6. Second detector law
7. Receiver gain
8. Video amplifier gain
9. CRT bias
10. Lobe width
11. Noise spectrum
12. Rotation rate of antenna
13. Beam focus
14. CRT anode voltages

Most of these have been studied on some kind of scope but only a few have been studied on more than one. There have been four major investigations made: one on an A-scope by Haeff (**16**), one on a PPI by Payne-Scott (**44**) and one using both an A-scope and a PPI by the M. I. T. group (**1**). The fourth is a continuing series of investigations on PPI's by several experimenters at the Systems Research Laboratory of the Johns Hopkins University. Both Haeff and Payne-Scott have attempted to summarize the effects of the significant parameters in equation form, Haeff using three parameters for the A-scope and Payne-Scott using four for the PPI. Payne-Scott's equation includes two additional constants based on "average" visual abilities of the observer; viz., one for brightness discrimination and one for area discrimination. Except for the M. I. T. group[5], which has published no theory to date but which is supposed to have developed one, other investigatiors have restricted themselves to empirical studies of inter-variable relationships. The possible combinations of 14 electrical parameters alone constitute a formidable array of possible experiments! It is doubtful whether it is worthwhile to attempt now an over-all mathematical expression for detectability for even one class of scopes, for the simple reason that a great many visual and psychological factors enter into detectability which are largely unknown. These will be discussed in the next section. If understood in the context of their limitations, equations can, of course, be useful in stimulating thought and new research.

Research on Visibility: Deflection-Modulated Scopes

Only two representative studies will be listed here; others exist, but are not as comprehensive. These two are engineering studies concerned mainly with electrical parameters.

Haeff of the Naval Research Laboratory (**16**) investigated pulse repetition frequency from 30 to 1670 pulses per second, pulse length from 2.5 to 20 microseconds, i-f bandwidth from .005 to 1.0 megacycles, and video bandwidth from .015 to 2.0 megacycles. He found that detectability varied directly with pulse length and with repetition frequency. There was an optimal band pass which was the reciprocal of the pulse length, although later Ashby *et al.* (**1**) showed that it may be

[5] A final report of the M.I.T. studies is presumably forthcoming as Volume 24 in the Radiation Laboratory series (**45**).

a little wider than this. With too narrow a band pass, signal strength is reduced and with too wide a band pass noise is admitted in excessive amounts. Either way, the signal-to-noise ratio is reduced. If noise is minimal or absent, presumably a wider band pass could be used. Video bandwidth had negligible effect.

In general, these results were confirmed and extended by Ashby *et al.* (**1**) at M. I. T., who also studied signal duration, focus, contrast, and detector law. Longer signals were detected better. Defocussing had almost no effect when parallel to the pip and very minor effects when perpendicular. No effect of trace brightness was evident unless the contrast was very near threshold. Whether a square law or linear detector was used made no difference, unless the signal was actually limited by the amplifier.

So far as is known, no current work is under way in any laboratory on A-scope detectability. Interest has shifted to the intensity-modulated scopes. There undoubtedly exist many questions yet to be answered, chiefly as to the peculiar visual abilities required in A-scope operation. Judgments of length of lines, of small separations between triangular areas, and of variations in linear extent have to be made in a limited time. The psychological processes involved are a rather different combination than is required in the usual visual experiment and have never been intensively analyzed.

Research on Visibility: Intensity-Modulated Scopes

Electrical Parameters

In an attempt to develop a theoretical integration, Miss Payne-Scott (**44**), of the Australian Council for Scientific and Industrial Research, carried out a whole series of experimental studies on electrical parameters. The series covered the following:

I-f Bandwidth.[6] With a constant pulse length, the bandwidth was varied and found to have an optimal value very little in excess of the reciprocal of the pulse length. This agrees with the A-scope data.

Pulse Length. Increasing pulse length improved detection very markedly with short pulses but less and less with longer pulses. The effect, of course, is partly dependent on the minimum spot diameter of the tube.

Pulse Repetition Frequency. With low frequencies the phosphor screen is excited intermittently and there are fewer echoes actually returned to it. Therefore, any increase in PRF improves the probability of detection up to the point where the scope background becomes uniform in appearance. Beyond this critical frequency, the threshold signal is said to be independent of PRF. The critical frequency is a function of antenna rotation rate. The lower the rotation rate, the lower the critical PRF. It is also a function of beamwidth.

Antenna Rotation Rate. If the combined PRF and rotation rate were sufficient to paint a uniform background, detectability was essentially independent of rotation rate. If the background was discontinuous and "grainy," the slower rotation rates were slightly advantageous, over a range of about 1 to 70 rpm.

Beamwidth. The beamwidth or lobe-width of the antenna largely determines the angular dimension (arc length) of the observed pip. The data showed that from 2° to 12°, detection improved as the $\frac{2}{3}$ power of beamwidth.

Video Bandwidth. In agreement with the previous A-scope data, no effect of video bandwidth was found.

Many of the same variables have been studied by others (see below) and most of the results have been confirmed. In constructing her equation for visibility, Payne-

[6] Although the author designates this parameter r-f bandwidth, she apparently means i-f bandwidth.

Scott chose four basic parameters: (a) the product of pulse length and i-f bandwidth, (b) the ratio of pulse length times sweep time of the CRT to spot diameter, (c) the ratio of antenna beamwidth to the angle subtended at the eye by the beamwidth, and (d) the ratio of PRF to critical PRF. With these parameters she was able to arrive at an equation which predicted fairly well the data taken in her own experiments, but failed in the case of large beamwidths. It also fails to include a factor for screen brightness, a critical variable, and of course the validity of extending the theory beyond present evidence is not known. The equation itself is not reproduced here, because of limitations of space, but the interested reader will be well repaid by examining it. There is no doubt that it represents the best approximation of an integration of radar visibility data that has yet been attempted.

In spite of the integrative value of the equation, it of necessity includes constants for the individual CRT and the individual observer. Of course, it does not even attempt to handle the supra-threshold brightnesses, multiple signals, or discriminability. Whether the equation is applicable to other phosphors and other tubes is not yet known.

Other Factors

The electrical parameters are only part of the story, although possibly the largest part. In the remaining pages a variety of factors influencing visibility will be discussed, even including some of the electrical parameters. The reason for the overlapping treatment derives from a psychological rather than an engineering orientation. Different electrical combinations often produce equivalent visual effects. The discussion follows a classification of *what is seen* on the scope rather than of what circuits are required for visibile operation. Payne-Scott's analysis is characteristically a radar-engineering approach; the following is a human-engineering one. Most of it is based on recent research by the writer and his colleagues at the Johns Hopkins laboratory.

Screen Brightness

Screen brightness is mainly controlled by the bias voltage on the control grid of the CRT; a secondary control is the gain of the video amplifier which amplifies noise as well as signal. There is a very definite optimal bias in a given tube, as is shown in Fig. 2. The legend to this figure explains the visibility measure used in this and subsequent graphs. The optimal bias in a typical 7BP7 tube was found to be about 5–6 volts less than cut-off, which corresponds to a brightness of the screen of about 0.1 foot-lambert in in the region adjacent to the pip. This bias is apparently less than that reported by Nottingham (**40**); the difference may be due to different experimental methods. At any rate, for objects of the size of the radar pip, the background brightness needs to be greater than this if visual sensitivity is to be maximal. (See Blackwell's data, **6.**) The level for maximal sensitivity varies with size of detail to be seen, but in general should be from 1 to 10 ft.-L on the scope. These levels are attainable, but only by reducing the responsiveness of the phosphor to incremental excitation, due to saturation effects. The brightness-voltage function is not linear at lower biases, i.e., brighter screens (cf. Fig. 3). The P7 phosphor does not, therefore, permit maximal use of visual sensitivity. Whether metal-backed screens (**12**) are better in this respect is not known; they have never been visually tested. The P10 phosphor also reveals an optimal bias (**15**).

Screen brightness is thus critical; and bias is perhaps the most important of all electrical parameters. Optimal bias is in turn dependent on video gain and on noise, for these can directly affect screen brightness. PRF and rotation rate of the antenna can also affect screen brightness. The effect of any circuit parameter ought first to be evaluated in relationship to brightness. If, for example, the tube bias is high (the scope, dim), then the addition of noise may actually improve the signal visibility because it simply adds brightness to the screen. (See Fig. 2.) Of course, the addition of certain frequencies of noise may impair visibility; for if the noise pips are the same size as the signal pips, the video gain is turned up; and the optimal bias should shift with gain. Garner and Hamburger (**15**) found this to be true.

Although bias and gain are critical factors, most investigators have ignored them. Payne-Scott (**44**), for example, merely says she set the bias at cut-off (extinction of elec-

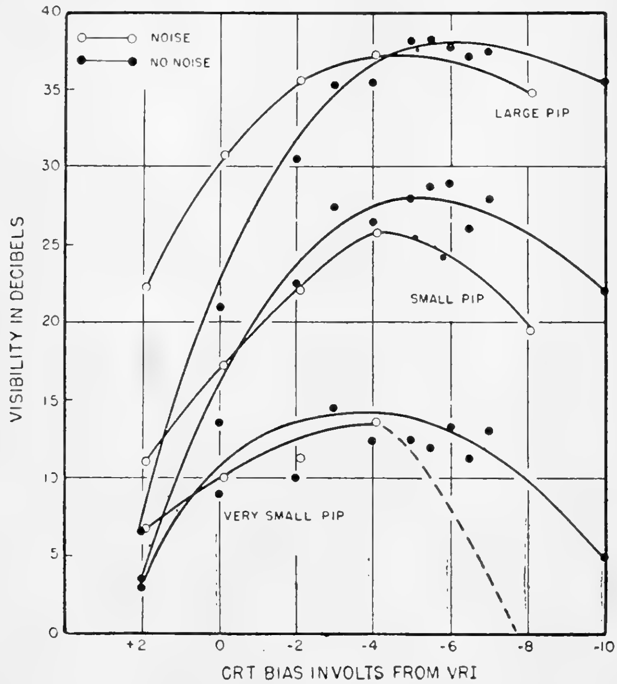

Fig. 2. Visibility, CRT bias, and video noise

Visibility is given in terms of the number of decibels by which a reference (8-volt) signal was attenuated when the pip was at the threshold of visibility. Increased visibility is therefore indicated by increased amounts of attenuation, i.e., by decreased signal voltages. Bias is a voltmeter reading taken from a visual reference as zero, i.e., the minimum visible sweep line with 10 rpm rotation rate and 667 PRF. Noise level is 0.85 volts rms (thermocouple type meter); noise random within limits of video band pass (100 cps to 5 MC) with i-f bandwidth greater than 10 MC. The large pip is 15 μs x 30°; the small pip, 2 μs x 2°; and the very small pip, $\frac{1}{2}$ μs x 1°. The optimal bias is about 5.5 volts less than the visual reference without noise and less than this with noise, but the brightness of the screen is probably the same for both optima. Being "grainy," the noise is detrimental to the discrimination of the very smallest pip. Broken line indicates invisibility at −8 volts.

probability of recognizing the signal is obviously reduced. If the tube bias is already at or below optimum, the addition of noise adds not only confusion but also unwanted phosphor excitation, and produces the equivalent of an even lower bias.

From this it must be evident that bias and video gain are interdependent variables, each tending to compensate for the other. Consequently, the bias should be higher when the tron beam) and then allowed the observer to adjust the gain as he pleased. No additional information is given.

Inasmuch as the optimal bias is dependent on the response curve of the phosphor and because phosphors change with use, there is no known electrical way of setting the optimal bias for a given tube by meters alone. Consequently, it has been suggested that a simple psychophysical method be employed by each operator every time he uses his scope (**55**); this could feasibly be done with the aid of an attenuated artificial signal which could be built into the radar indicator. The suggestion has not thus far been acted upon.

A very interesting implication of the foregoing analysis is that a signal which reduces

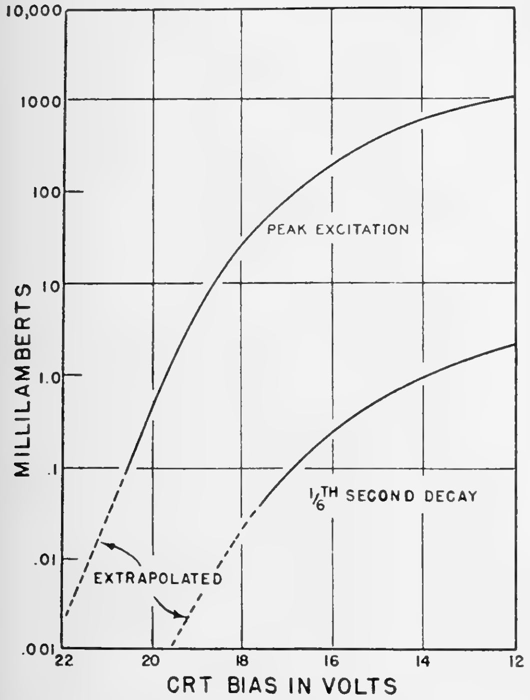

Fig. 3. Screen brightness and CRT bias

A brightness-voltage function for a typical 7BP7 tube, measured by a photomultiplier tube. Actual voltage values vary with the tube and with its age. The visual reference for this tube was slightly less than 22 volts, optimal visibility was at 17 volts, or about 0.1 ml for $\frac{1}{6}$th second decay. Reprinted from Williams and Bartlett (**54**).

screen excitation should present a more visible pip than one which *increases* it, *provided the bias is optimal or less.* Because at medium and high excitation levels the phosphor screen becomes less and less responsive to new excitation (Fig. 3), the signal is, in effect, "clipped" by the screen. Therefore, a given signal *voltage* quantity will produce less increment than decrement of brightness; the result will be a greater visual *contrast* in the decremental pip. To test this hypothesis, Harriman and Williams (**22**) reversed the polarity of the signal lead-in wires to the CRT: instead of putting the signal on the control grid, they put it on the cathode. The result was an intensity modulation in the opposite direction: the pip appeared darker, rather than lighter than the background. Threshold measurements made with the two types of pip are graphically shown in Fig. 4. Two things are to be noted: (a) with brighter scopes, the ultimate visibility of the "dark" pip is better than for the "bright" pip, and (b) the optimal bias is lower (i.e., screen brightness higher) for the "dark" pip, at least for small pips. The latter means that the screen can be operated at a brightness of about 0.5 ft.-L for optimal detection rather than at the 0.1 ft-.L necessary for the bright pip. Very likely, some part of the increased visibility of the dark pip is due to the higher adaptation level of the eye.

It is to be remembered, of course, that the bias which is optimal for *small, weak* signals *may* not be so for other operational purposes. For example, spot size increases in some unpredictable manner with beam current; in fact, at very high beam currents it may increase very rapidly and even exhibit "blooming," a peculiar phenomenon of light spreading which apparently is not well understood (**2**). Whether, short of blooming, the increased spot size at moderately higher currents (brighter scopes) is sufficient to impair resolving power, is not yet known.

Pip Size

The size of the perceived pip is a crucial variable in its visibility. Radar pips may be as short in duration as $\frac{1}{2}$ microsecond or as long as several microseconds; pulse duration determines the radial dimension on a PPI. Pips may be as narrow in beamwidth as one or two degrees or as wide as 30 or 40 degrees, depending on the lobe pattern of the antenna transmission. The effective beamwidth of a pip from a single target may vary considerably as the distance from the antenna, because lobe patterns typically are rather irregularly "pear-shaped" and will cause an echo to show up as wider at medium than at short or long distances. Beamwidth in the present context will refer only to actual beamwidth of the pip on the scope, i.e., the length of the arc segment (cf. Fig. 1). In linear measure, the angular width of a pip varies as its distance from the center of the scope. In angular subtense at the observer's eye, a very small pip $\frac{1}{2}$ microsecond x 1 degree would, if seen from 12 inches at half-radius on a 6-inch scope and a 20-mile range scale, measure 1′48″ x 7′31″. An "average" size pip might be 2 microseconds x 10 degrees, which would subtend an area about 7′6″ x 1°15′. Pips, on the whole, are rather small objects, and therefore require maximal viewing conditions.

The effect on visibility of changing pulse length and beamwidth can be seen in Figs. 5 and 6. These data were taken on a noise-free scope but very similar results have been obtained with noise. The total range of pip sizes is very large compared to those encountered in operational conditions. Except in the case of the dim screen, a different law is seen to hold for beamwidth than for pulse length. For small pips, visibility increases more rapidly with pulse length than with beamwidth, even when the latter are equated in retinal subtense. This effect is presumably due to the time difference in rotary presentation: all points on a radial

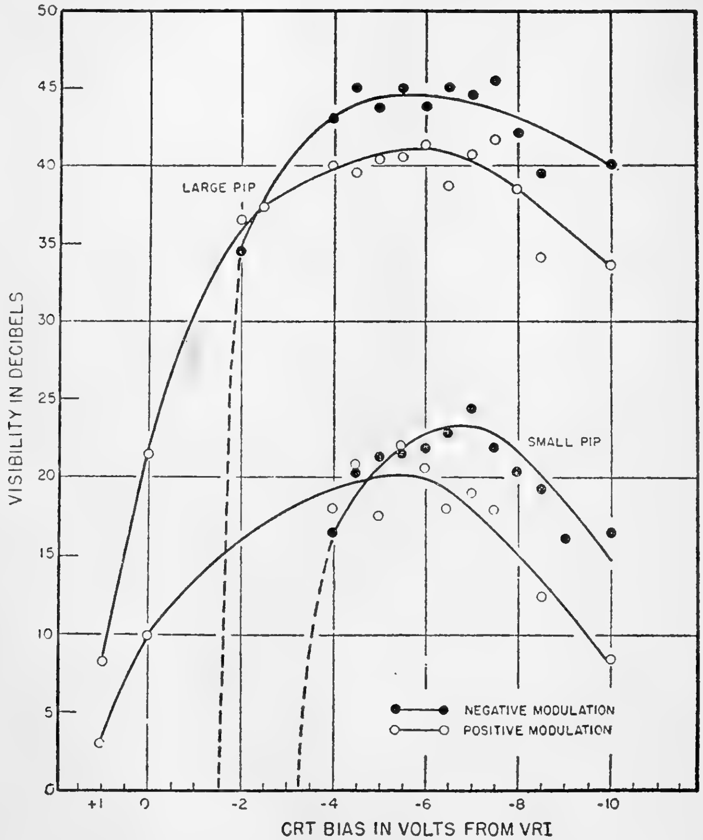

Fig. 4. Positive versus negative modulation of beam

Positive modulation is normal, producing a brighter-than-background pip. Negative modulation produces a pip darker than its background. Large pip is $7\frac{1}{2}$ μs x 10°; small pip, 1 μs x 1°. No video noise. Otherwise conditions same as in Fig. 2. Broken lines indicate invisibility.

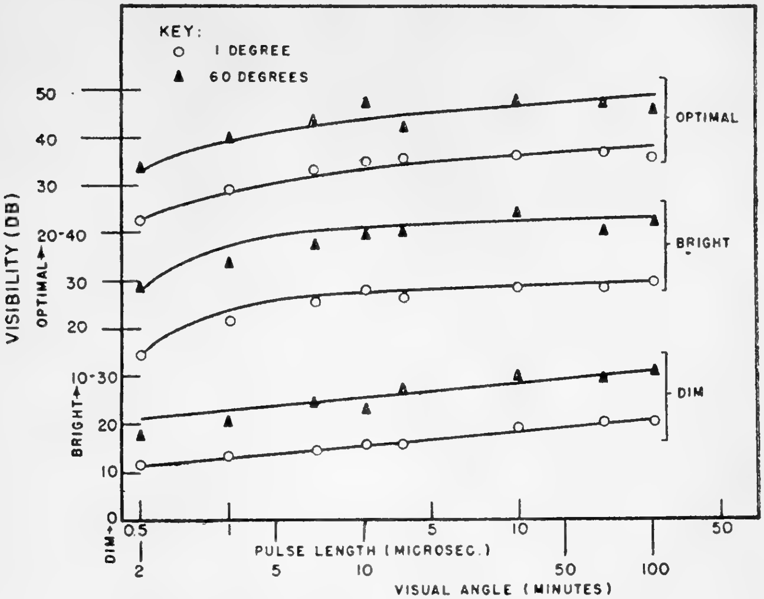

Fig. 5. Visibility as a function of pulse length

Data are shown only for two beamwidths (1° and 60°) and for three bias levels. *Dim* is the visual reference bias and is, effectively, about 0.0001 ft.-L; *optimal* is 5.5 volts less, about 0.1 ft.-L; *bright* is 10 volts less than the reference, about 2 ft.-L. No video noise. Otherwise conditions are same as in Fig. 2. Reprinted from Bartlett, Williams and Hanes (5).

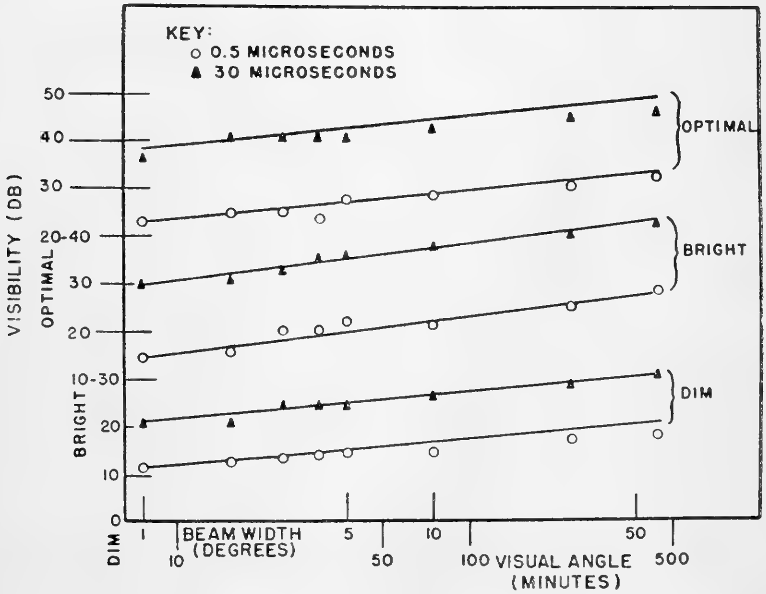

Fig. 6. Visibility as a function of beamwidth

Conditions same as for Fig. 5; reprinted from Bartlett, Williams and Hanes (5).

dimension are exposed simultaneously, whereas points on the angular dimension (azimuth) are exposed successively. Radial sizes, then, correspond more to those in optical presentations; angular sizes, above a few degrees, are more of a gradual "uncovering" exposure: as the cumulative area gets bigger there is an increasing probability of a unit area's being looked at.

Other Factors Influencing Pip Size

Several factors influence effective pip size besides *pulse length* and *beamwidth*. Some of these are as follows.

Sweep Time or Range Scale. On a radar set, the range scale is the control for sweep transit time, i.e., the time required for the deflection of the electron beam from the center to the periphery of the scope. Since scope radius is constant, a one microsecond pulse will occupy a greater proportion of the radial sweep on a short distance range scale than on a long. It will appear, in fact, 10 times greater in radial length on a 20-mile scale than on a 200-mile scale. This relation has been investigated and the expected effect on visibility obtained (**4**).

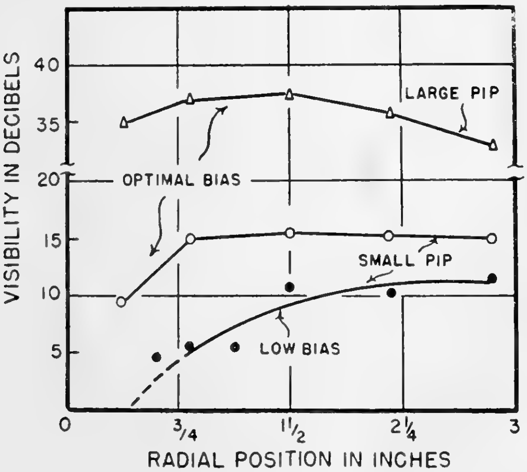

Fig. 7. Visibility and range

Range of target is defined by position on 3-inch radius. Data are for a typical 7BP7 tube, no noise. Large pip is 3 μs x 10°; small pip, $\frac{1}{2}$ μs x 1°. Low bias is 10 volts less (brighter) than visual reference; optimal bias is 5.5 volts less.

Range Position. As a pip closes in range it moves closer to the center of the PPI; and though its angular beamwidth remains constant, the same degree of modulation is placed on the electron beam at all range positions. This means that, although the pip is becoming smaller at close range, it is also becoming brighter, unless limited by phosphor saturation. Experimental tests (cf. next section) have not consistently shown any effect of range. Data sometimes show a drop-off in visibility at the extreme periphery or very near the center, but these effects usually may be attributed to poor phosphor qualities, i.e., "burning" at the center and thinness at the edge. There is, however, good reason to believe that visual size may be a factor, under certain conditions. The argument is as follows: there is an approximate compensation of area loss by brightness, i.e., the $A \times I = C$ (Ricco's law) statement is roughly true for small objects. Apparently what happens in the usual PPI scope is that the increasing brightness of the pip, as it approaches the center, approximately compensates for the decreasing beamwidth. This is to be expected under favorable brightness conditions, but if the phosphor is nearly saturated the brightness will not be able to increase proportionately, the total light increment will be reduced, and visibility will be more impaired. This effect should be greatest with very small pips. An experiment to test this expectation more or less confirmed it. The data are shown in Fig. 7.

Distance of Observer from Scope. As the distance of the eye from the scope increases, the retinal subtense of the pip decreases and therefore visibility is expected to become poorer. This is true, as shown by Bartlett and Williams (**4**), for noise-free scopes but there was no advantage of the closer view when the scope was so heavily cluttered by large noise pips as to require considerable discrimination among them. The maximal advantage of a 6-inch over a 24-inch view was 8 db.

Scope Size. Within wide limits, the scope diameter will probably not affect the visibility of large pips, but for very small and weak pips, the larger scope should furnish slightly better visibility, in accordance with the laws governing size effects. No direct experimental comparison of scope sizes under these conditions, with every other feature of the CRT held constant, has yet been reported.

Shape and Contour Variables

Pulse Shape

The pulse transmitted and received ordinarily is rectangular. Deviations from rectilinearity can be expected if the receiver band pass is too narrow; this results in reduction of pulse energy and lower visibility. Even with energy equated, two pips differing in temporal distribution of energy may be unequally visible because of different contour gradients. That is, a pip which has no sharply defined contours in the radial dimension would probably be much less visible than one with sharp borders. This problem has not yet been investigated on scopes, but needs to be.

Azimuth Gating of Pip

The angular or beamwidth dimension of the ordinary radar pip is of irregular brightness; usually the pip is brighter in the center than at the leading or trailing edges, thus producing a roughly triangular rather than rectangular distribution. Plans have been made by the writer to investigate the effects of azimuth gating methods, but the equipment has not yet been perfected. The experiments are needed so that we may know how to distribute a given quantity of signal energy most efficiently. The sharp, rectangular pip pictured in Fig. 1 is definitely not representative of operationally encountered pips.

Focus

In lieu of precise information on the variation of energy in the separate dimensions of the pip, the effect of defocussing or blurring it may be of interest. Focus is established by a special focus coil, whose voltage can be varied over a wide range—in a typical 7BP7 tube, from 168 to 322 volts. The best focus is between these extremes. In an unpublished study by the writer, a trained operator set the "best focus" of the sweep line at each of several bias levels. The threshold intensity was then determined for a typical pip ($\frac{1}{2}$ microsecond x 1°) at the "best focus" and at each of six other focus coil voltages (three greater and three less) from 168 to 322 volts. The results are shown in Table I.

TABLE I

FOCUS, BIAS, AND VISIBILITY

Columns to the right show data for different bias voltages. These vary from high to low in order to effect scope brightnesses from dim to bright. Optimum detecting bias is between −4v. and −6v.

"Best focus" means the sharpest, least blurred sweep line as judged by direct visual inspection, no pip being present. Whether to focus on sweep line or on pip makes little difference in *visibility*, as shown by line *c*. On the other hand, extreme de-focussing does have a serious effect, as shown in line *f*, which varies with scope brightness.

	VRI (Dim)	−2v.	−4v.	−6v.	−8v. (Bright)
(a) "Best focus" voltage set by operator	219	203	209	211	215
(b) Focus voltage giving best visibility of pip	202	180	209	196	215
(c) Visibility advantage of (b) over (a)	1 db	1 db	none	1 db	none
(d) Visibility score at (b) in db signal voltage	31	36	42.5	44	42
(e) Visibility score with extreme defocussing	30	32.5	34.5	31	24.5
(f) Visibility loss due to defocussing (d) minus (e)	1	3.5	8	13	17.5

In defocussing, virtually no energy is lost; it is simply spread out more, so that the pip becomes very fuzzy. Very interestingly, this profound change appears to have no significant effect on a dim scope, where the sweep line itself is barely visible (VRI or Visual Reference Intensity), but has a marked effect on a bright scope. Although the loss at the brighter levels can be attributed to contour mutilation, the size of the effect is magnified somewhat by the db unit, whose light-output equivalent varies with grid bias. The explanation of the major effect no doubt lies in the eye; it integrates total light in an area (A × I = C) at low luminances, but requires brightness contrast more than area at high luminances and requires it in *adjacent retinal regions* (contour or border effect). (See Lamar, Hecht, Shlaer, and Hendley, **27.**) The defocussed pip on a bright background evidently does not supply the needed brightness contrast. More detailed and analytic study of this matter is desperately needed.

Orientation Factors

Range of Target

With ordinary-sized pips and normal operating conditions there appears to be no marked effect of range position on visibility, as stated previously. This presumably is due to an approximate compensation of brightness and area. Unfortunately, no direct physical measure of luminosity of the pip at various ranges has ever been made. This needs to be done. From visibility studies we can conclude that the middle ranges are about equally favorable for pip detection and that there may or may not be a fall-off in visibility near the center or near the periphery depending on size and brightness of the pip and the condition of the scope.

Bearing Position

In looking at a scope of the PPI type, will it make any difference whether the pip appears at north, east, south, or west? The answer is, apparently not, except under certain peculiar circumstances. In the course of thousands of observations of pips on PPI's during experiments at Johns Hopkins, no effect of bearing position has ever been noted, with one exception reported by Garner and Hamburger (**15**). This special case appears to be due to a glare factor uniquely associated with the large projection PPI.

Sweep Direction

An unpublished study by the writer showed no significant difference between clockwise and counter-clockwise direction of movement of the sweep line.

Scope Position

Similarly, unpublished data of the writer's showed no effect of scope position provided the observer's normal line of sight was maintained.

Angle of Viewing Scope

Tilting the whole scope is quite another matter. Considering scope curvature, it is by no means obvious how much tilt would be required to impair visibility. Therefore, another experiment was undertaken by the writer in which visibility thresholds were tested with the scope tilted at 90, 80, 70, 60, 45, and 30°, the observer's head remaining fixed. The threshold for a single pip was unimpaired from 90° through 60° of viewing angle. At 45° there was a drop-off of 3 to 4 db and a further drop-off of about 3 db at 30°. No thresholds were taken at angles less than 30° because it was evident that the observer would soon be unable to see anything. Incidentally, the same trend was shown both for conditions of darkness and with a very slight room illumination (about 0.1 ft.-c) insufficient to produce glare on the scope face. Unquestionably, viewing angle would become a more critical factor in the presence of glare. (For these tests an uncovered 7BP7 cathode-ray tube was used; the observers looked directly at the face of the tube without intervening cover glasses or filters.)

Decay Time (*Signal Age*)

Sweet and Bartlett (**48**) measured visibility thresholds for a single pip at several periods in its decay history. They did this by an ingenious device, which is essentially a rotating tachistoscope fitted over a standard scope. A slit-like aperture rotated with the sweep line, either directly over it or lagging it by a predetermined interval. The rate of decay of scope brightness during afterglow varies (roughly inversely) with the intensity of the original excitation. Both pip and background decay together, of course, but the pip, being brighter, decays at a slightly different rate. The optimal decay interval for visibility was found to depend on screen brightness, so four bias levels were tested. In general, visibility decreased with time but the decrease was more pronounced on bright scopes. On a dim scope the pip is best seen when directly on the sweep line (zero decay), but on brighter scopes it is seen better just *after* the sweep line has passed (.04 second decay better than zero or .5 seconds; intervening ages not tested). The visible persistence of signals was also observed and found to vary from zero to more than three minutes, depending, of course, on signal strength and bias.

This experiment can be regarded as a pioneer. It offers a technique for the precise measurement of phosphor variables which might profitably be employed in a phosphor laboratory. Visibility-decay curves should be determined at least for the P12 and P14 phosphors. Of particular value would be the inclusion of pip size as a parameter.

Illumination Variables

The *intensity* of external illumination is a critical factor for two reasons: (1) it can reduce the effective contrast of the pip by veiling glare on the glass surface of the scope, and (2) it can alter the sensitivity of the eye by changing its retinal adaptation. Studies have been made of both effects.

Intensity of Illumination

Ambient illumination has been varied from very nearly zero to several foot candles by Williams (**51**) and by Williams and Hanes (**56**). Mazda light, moderately diffused, is not detrimental to visibility on a PPI if kept at or below the level of screen brightness, at least when the latter is about 0.1 ft.-L.; it may safely exceed the screen by a log unit or slightly more, provided the scope itself is hooded or shielded (**51**). Reproduced here as Fig. 8 are representative data taken not only with Mazda but also with other colors of illumination. It is fairly clear that light very much in excess of the scope brightness is damaging, no matter what the color. Craik and MacPherson (**10**) also found an intermediate optimal intensity, but did not specify its relation to scope brightness, which was not measured. So far as is known these are the only studies ever done using PPI scopes. It had previously been shown that illumination is much less critical for A-scopes provided its pips are fairly bright (**34**). There is an enormous literature based on test objects other than radar pips. A good review of illumination intensity is given by Crouch (**11**).

Adaptation of the Retina. Adaptation apparently is not the limiting factor in ambient light level, except in the special case of darkness. Hanes and Williams (**19**) found that visibility of the pip was scarcely affected by adaptation of the eye to levels from nearly zero to 8 ft.-c., *provided the scope itself was operated at its optimum of about* 0.1 *ft.-L.* In fact, adaptation to darkness may be very slightly damaging because of the short period of light adaptation required to come up to 0.1 ft.-L. Of course, if the scope is very dim, the effects of high light adaptation can be sizeable. The instantaneous threshold is raised and detection time is increased. (This is another reason why scopes should be operated at brightnesses as high as is consistent with visibility.) In general, the rule holds that the impairment of visibility is proportional to the difference between the adapting and the test illuminations, but small differences are insignificant. An excess of a few foot candles is unnoticeable in its effect on visual adaptation *per se*, but has a marked effect when shining on the scope face *during* observation.

Color of Illumination

The color of illumination has been alluded to above, and relevant data from Williams and Hanes (**56**) are given in Fig. 8. At very low illuminations color is certainly insignificant. At moderate levels, near that of the scope, and at higher levels, there appear to be very small but rather consistent differences in favor of the shorter wavelengths. Whether any practical advantage could be gained from colored illumination in radar operating rooms is dubious; it would at best well be sizeable "emotional" effects of color which might affect extended work if not thresholds. Qualitative observations of pips under colored illumination suggest that there may also be some peculiar perceptual-constancy effects. For further work on color, the reader is referred to the Eastman-Kodak report (**58**) and to Ferree and Rand (**13**).

Filters

Because of practical interest in the matter it may be mentioned that unpublished data

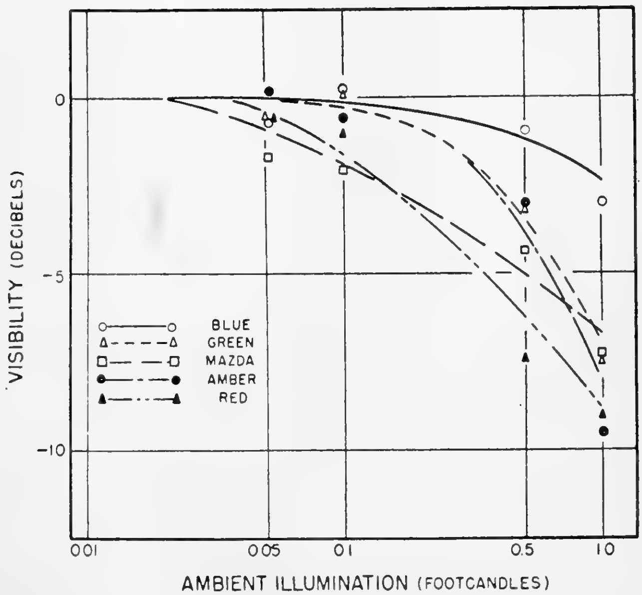

Fig. 8. Visibility and ambient illumination

Relative loss in visibility as each illuminant is increased from 0.02 ft-c., taken as a reference. Actually, any of the illuminants was ½ to 1 db better than darkness in intensities up to about 0.1 ft.-c. Scope brightness is about 0.1 ft.-L. (optimal bias). Conditions: 7BP7 tube; 2 μs x 2° pip; no video noise.

be very small. Certainly one would expect no differential effect *on thresholds* merely of colored reflecting surfaces near by, such as wall paint or paper. As an academic question, the effect of color of illumination is more interesting because it still is largely an enigma. Most experiments have shown either no effect or no consistent effects of color unless the visual task itself is one of discriminating colors, as in making color-matches of dyes and fabrics. More work in this area certainly is needed, for there may of the writer's show no very sizeable decrement in visibility when a sheet of ordinary plexiglass is put over a PPI scope. With negligible ambient illumination, the mean loss with clear plexiglass was 1.4 db and the mean loss with an amber plexiglass, the kind used for Navy scopes, was 2.0 db. The loss was slightly less when ambient light was increased.

Treated filters, i.e., plexiglass with a fluoride coating evaporated on the surface, were also tested by comparing them with untreated plexiglass (also untreated vs. treated halves of the same pane). With negligible external illumination, there was no advantage of the treated filter. There was a definite though slight advantage of the treated panes when there was a significant amount of room illumination; signals required 0.5 db to 1 db less strength to be seen behind treated than untreated panes. This is easily attributable to reduction in specular reflection. Both types of pane caused some loss in visibility. For example, in the presence of 20 ft.-c of illumination, the untreated plexiglass was 1.5 db worse than no filter at all but the treated plexiglass was only 1 db worse. Visibility tests of coatings on the bulb itself would seem desirable.

Uniformity of Background

Excitation of a PPI scope is a type of *discrete* stimulation: pulses are separated in time. Successive pulsations do not ordinarily fall on quite the same spot on the phosphor screen because the time sweep (radial or range deflection) keeps "painting" them on at gradually increasing ranges and the rotation of the antenna continually moves the sweep around the circle of the scope. Obviously, if the pulses are few and far between and if the rotation speed is great, the rotating sweep will actually miss many spots on the phosphor. "Silent" gaps of no excitation appear and the observer sees only an intermittent flickering and "spoking" of lights. The intermittence decreases as the repetition frequency of the pulsing increases and as the rotation becomes slower. To a lesser extent it is also dependent on pulse length; and its visual appearance, of course, depends on brightness as well as on construction features of the phosphor screen, etc. The critical construction feature is the "build-up" ratio of the screen which is a function of its decay law; luminous build-up occurs when successive stimulations overlap, as they do under the usual firing and deflection rates. With a given screen, and a constant pulse length and bias, though, the background will cease flickering and appear uniform above some critical ratio of PRF and rotation rate, *for a particular observer*. (Here is an application of flicker data to radar.) A unique variable in the equation is noise. Noise is by definition a random population of pulses which, of course, is no longer entirely random as to duration after passing through frequency filters in the receiver (r-f, i-f and video bandpasses). Because of its random character, noise can be mathematically evaluated only on statistical assumptions. It is here that the elaborate statistical theories be-become important (**44, 46, 50, 59**).

From a perceptual standpoint uniformity of background may be a kind of intervening variable, i.e., a requisite condition which is independent of its manner of production. Such an hypothesis needs two kinds of data for confirmation: direct, subjective judgments of the "quality" of uniformity—a spatial analogy of critical fusion frequency; and pip visibility thresholds. Both should be dependent on the electrical parameters in roughly the same way.

Subjective Judgments

Attempts in the writer's laboratory to judge the threshold of afterglow uniformity have been rather disappointing because of the difficulty of maintaining a constant criterion of judgment in the face of so heterogeneous a stimulus display. Nevertheless, the judgments do correlate roughly with PRF and rotation speed in the expected manner. A PRF of 50 or 100 pulses per second requires a very slow rotation speed (3 rpm) to produce uniformity, whereas a PRF of 300–400 produces the effect with as much as 30 rpm. Judgments of the sweep line are not quite so difficult. Even here, though, frequencies low enough to appear intermittent are disturbing. The bias required to produce the minimum sweep is lower for low PRF's than for high and the PRF for low antenna speeds is smaller than for high. These observations are at least consistent with expectations.

Visibility of Pips

The threshold values of a pip as a function of PRF and antenna rotation speed have been investigated by Payne-Scott (**44**) and the results have been mentioned above. Scott says. But this is true only for very bright scopes or for one whose bias is adjusted to PRF to maintain a constant scope brightness. The change-over frequency is not necessarily the threshold of background uniformity; this too depends on bias. With dim scopes, visibility improves with PRF even *after* the scope is uniform because more

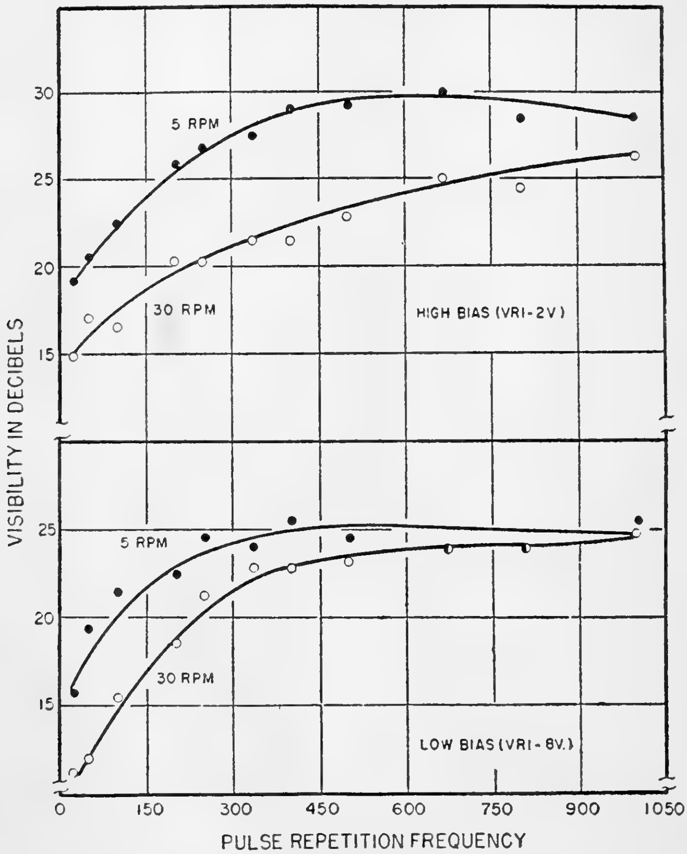

Fig. 9. Visibility and pulse repetition frequency
The effect of PRF depends both on bias and rotation rate. The lower curves are for a bright, nearly saturated screen; the upper, for a dim screen. Conditions: 7BP7 tube; no video noise; 2 μs x 2° pip.

No one has yet worked out in detail all of the interrelations among the parameters although Payne-Scott's theory is the best approximation. Suffice here to illustrate a few relations.

PRF and Rotation Rate. Data of Williams (**52**) show that visibility may improve with increasing PRF up to a point and then be essentially constant, as Payne- pulses per second can still add a favorable brightening to the screen. With very high PRF and very low bias, rotation speed can be removed as a factor in pip visibility. Otherwise it is inversely related to visibility. (See Fig. 9.) Equipment limitations have thus far prevented exploration of the very high PRF's and rotation speeds.

PRF and Bias. Part of the improvement

sometimes accompanying higher PRF's is due to screen brightening which is, in visual effect, equivalent to reducing bias. The writer made a simple test of this by varying PRF from 25 to 1000 pps, in 11 steps, adjusting the bias at each PRF so as to keep the sweep line just barely visible. The compensation required over this range was 1.3 volts of bias; that is, the screen had to be brighter, by an amount produced by 1.3 volts of bias reduction, with a PRF of 25 than with a PRF of 1000. However, when bias was thus adjusted for each of the 11 PRF's, the threshold intensity of a medium-sized pip was absolutely constant, within the error of measurement. In short, visibility was *independent* of PRF *per se*; it was dependent on the brightness consequences of PRF. This is precisely what we would expect when using dim scopes. Likewise, if the scope is already too bright, additional brightness caused by increasing PRF may be unfavorable rather than favorable. (See Fig. 9.)

Video Noise. The whole problem of noise (PPI "snow" and A-scope "grass") deserves better treatment than can be given here. The mathematics of noise reception and filtering has not been completely worked out, especially in relationship to video indicators, though some treatises are available on the subject (**46, 50**). From a perceptual standpoint, noise tends to reduce homogeneity of the screen. It also adds brightness. High frequency noise may merely make the PPI screen look "grainy"; low frequencies may make it appear "blobby" and patchy. Large beamwidth targets, appearing as arc line pips, may readily be distinguished from most noise blobs; but extremely small pips may resemble the shape of a noise pulse so much as to be distinguishable only on the basis of brightness or recurrence. A really satisfactory experimental video noise, variable in specturm and intensity, is not known to have been employed in any experiments to date. The data which come closest to satisfying the requirements of spectrum analysis are those referred to above on i-f bandwidth. Evidence on the brightening effect of noise can be gleaned from Fig. 2. The optimum of the visibility-bias curve appears to be shifted about 1.5 volts to 2 volts; this particular noise (random; wide band pass) was lighting up the screen in an amount equivalent to 1.5 volts bias; being fine-grain noise, it did not reduce the ultimate visibility of the pip. In contradiction to our earlier conclusion (**55**), based on tests of non-random noise, the presence of *random noise definitely lowers the optimal bias*, probably in proportion to its rms value.

Optically simulated noise was used in one experiment by Hopkinson (**24**). He took photographs of operational radar scopes and projected a simulated pip from a light source on to the photographic representation. Only preliminary data are reported, but these show that the contrast threshold is the same function of background brightness, with or without noise.

The clutter that appears on operating radar scopes is not all noise; much of it consists of echoes from the sea nearby, from clouds, birds, and objects of no operational significance. This clutter is a serious hazard to detection; but being difficult to reproduce in the laboratory, it has been largely ignored by the researchers.

Visual Scanning and Search Methods

Search Area

As a radar operator searches a PPI scope for new targets, he must continually keep his eyes roving over the scope, usually following the sweep line, hunting for new pips. Is a whole scope less efficiently scanned than a part of it? Unpublished data of the writer's (**57**) showed surprisingly little effect on visibility of variations in fractional area of the scope to be searched. In a comparison of $\frac{1}{8}$ with the entire 6-inch scope, appearance thresholds were only about two db better for the smaller area. This perhaps ought not to be surprising in view of the width of the visual field (cf. Fig. 10 below) and the speed of eye movements.

Retinal Distance from Line of Sight

Although in general it is known that in good light, foveal vision is better than peripheral, the effect of eccentric vision on radar detection had to be determined directly on radar scopes. Data taken from Williams' study (**53**) are shown in Figs. 10 and 11. It is evident that for the smallest pip available, visibility steadily declines as the pip is moved away from the point of visual fixation and that there is no real area of equal visibility. For a larger pip there is a small area of equal visibility, although it is somewhat less than shown because of the fact that the distance from the fixation point to the pip was measured from center to center. Considering, however, that the smallest pip subtended only about 1′48″ at the eye and the largest pip only six times this, the extension of the pip itself into the fixation area could account for but a fraction of the equal visibility band. It is also interesting to note in Fig. 10 that the slopes of the curves are different for different background brightnesses. As might be expected, the steepest slope is with the brightest screen and the least slope with the dimmest. These data confirm, in general, those of Craik and Macpherson's (**10**) similar study, but differ in yielding a slightly larger effect. Whether the difference is due to background brightness is not known, for no mention of this factor is made by them. There is an interesting implication of these data for radar practice. For example, it had been previously noted by the author that if an attempt is made to trick an operator by getting

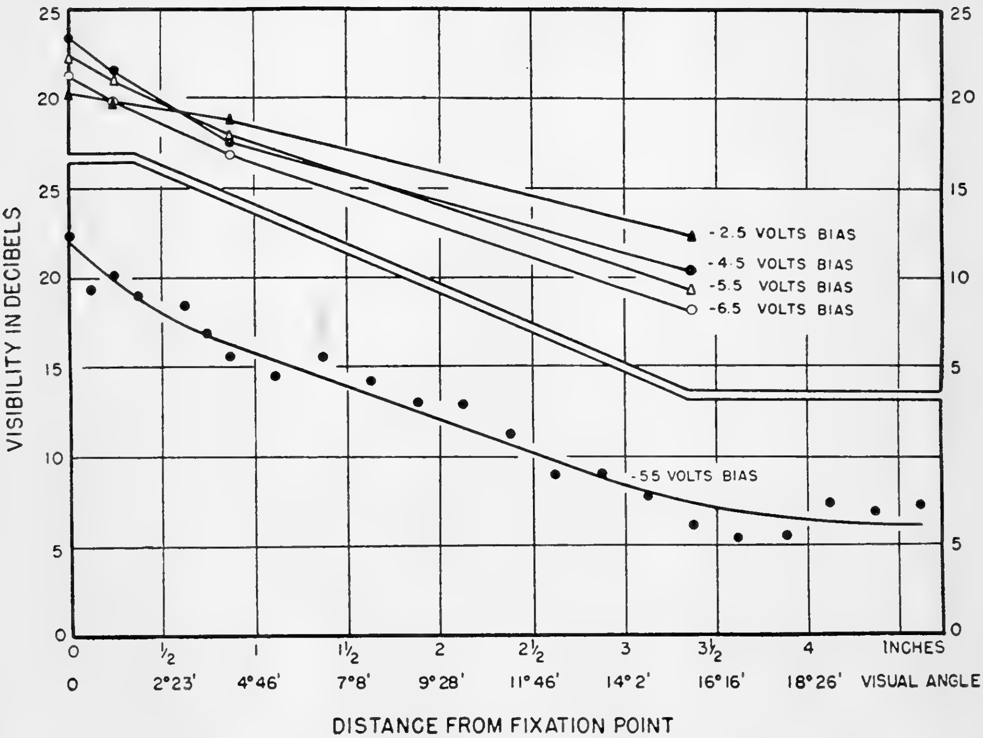

Fig. 10. Visibility at different retinal positions

Binocular fixation was at zero. A 1 μs x 1° pip (retinal subtense, about $3\frac{1}{2}'$ x $7\frac{1}{2}'$) was shown at various distances in a horizontal plane... The rise at 4 inches is probably an artifact of method. No video noise. The data from the upper and lower sections were taken in independent experiments.

him to look at one side of the scope and then "sneaking in" a pip at the opposite side of the scope, he might not notice the presence of the pip until it was as much as 6 or 8 db brighter than threshold. The experiments quantify this effect. The maximum loss to be expected from attending to the

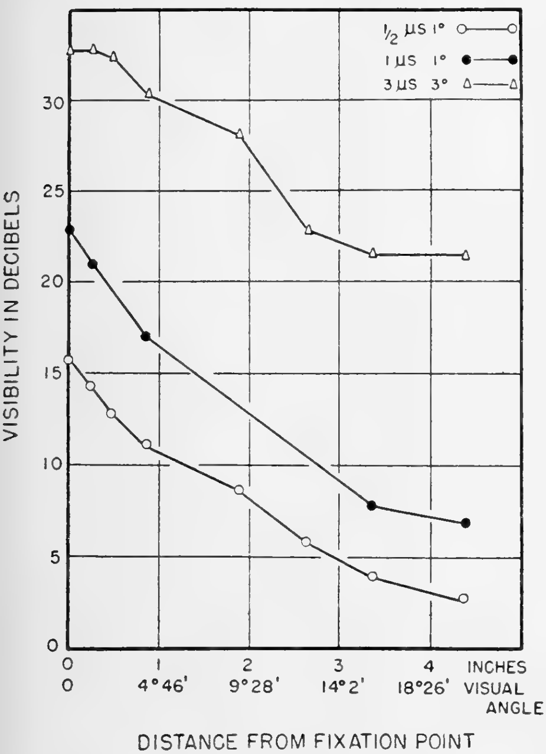

Fig. 11. Retinal position and pip size

The largest pip, 3 μs x 3°, subtends about 10½′ x 22½′ at the eye. Distances from fixation point to pip were measured from center to center. No video noise; optimal bias (−5.5 volts from VRI).

wrong part of the scope is about 16 db; average losses would be much less.

Search Time

The time required to find a target in a given area seems to be a topic not much investigated. In radar work it is a very practical question. Therefore, Harriman (**21**) carried out an experiment on the matter, the data of which are shown in Fig. 12. Although only two observers were used, each point on the graph is a median of at least 35 observations per observer. The interesting thing is the steady increase in search time as the pip gets weaker and weaker. At least within the range from threshold to 10 db above threshold, search time is a continuous monotonic function of pip brightness (signal strength). This would not be notable were it not for the fact that a 10 db range in such signals is at least four and

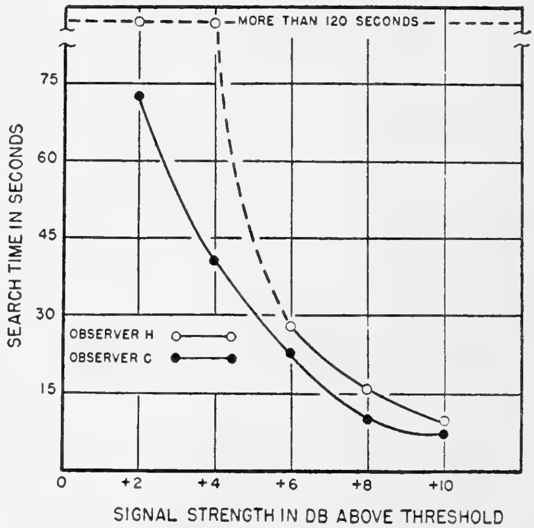

Fig. 12. Search time and signal strength

Search time was limited to two minutes. Bias, optimal; no noise. The rotation rate of 30 rpm means that the pip was re-excited every two seconds.

probably five times the interval of uncertainty which, of course, always brackets any threshold. Due to the inherent variability of threshold data of this kind, which normally fluctuate from one to two db, the near-threshold portion of the curve is essentially indeterminate.

Observer's Confidence and Criteria of Judgment

Different observers no doubt vary from one to another in the criterion used for visibility. Some may report a pip when they think they see "something." Others may wait until they recognize a characteristic form. It would be interesting to know just how much practical difference this would make in radar work. Though not obtained on real radar indicators, Hopkinson's data (**24**) are easily the most relevant. He had his observer use four different criteria of visibility: O, just visible; C, just recognizable and easily visible; B, just visible with some discomfort, enough judged to cause fatigue in extensive work; and A, two traces just recognizable as double. Although these sound like peculiar psychological categories, they probably made sense to the observer who used them. At any rate, most of the visual functions investigated followed almost precisely the same form no matter what the criterion; that is, the curves were all nearly parallel. Judging by his published graphs (no raw data were given), it appears that the largest difference, that is, between the just visible (O) and the easily visible (C) is on the order of a log unit-and-a-half of contrast. Another kind of data relevant to this problem are those comparing the method of average error with the method of limits, that is, the adjustment threshold compared with the appearance threshold. At various times the writer has made this comparison with the result that the latter are greater than the former by approximately two db, although the size of the difference may increase with low PRF or other special conditions.

It is commonly known among radar operators that, depending on conditions, a person may not report a pip seen only once. In short, the operator tries to report targets, not pips. When pips fluctuate and recur only intermittently, what does it take to make the operator sure that he can report a target? Space limits discussion of this interesting matter, but it is probably evident that it raises many psychological problems of great import. Unfortunately, no scientific answers can be given at this time.

Prolonged Search and Alertness

Attitude and Interests of the Operator

The total number of targets identified by a radar operator is probably more dependent upon the operator's alertness and attention to his radar scope than to any other single factor. The reason is quite obvious: if the operator is not looking at his scope he cannot see anything on it. Informal operational accounts and comments of military officials agree that the operator often is not looking at his scope even when he presumably knows that he ought to be. If this is the case, a study of the factors causing operator inattention ought to receive first priority in research. That it has not, is no doubt due to the psychologists' inability to find any practicable measures of such intangible entities as attention and alertness. Various schemes and devices have been suggested for keeping an operator on the alert. Occasional false "alerting" pips, "dummy" targets, more motivation by more pay or higher military status, associated auditory signals, automatic alerting bells triggered by physiological concomitants of drowsiness, and many other schemes have been mentioned. None has proved itself as a practical solution. Many a scheme might solve part of the problem, yet fail to turn a wide-awake operator's eyes from a magazine back to the radar scope.

Length of Duty Period

Another associated problem is the familiar one known as "fatigue." It is now agreed that almost nobody knows much about fatigue or boredom except from his own subjective experience. Consequently, measures of performance decrement due to fatigue tend to be unreliable and even contradictory. There are no known data which show any effects of fatigue on visibility, although there are two studies on efficiency of visual work. The first is OSRD Report No. 3334 (**35**), which showed a definite loss in time of efficiency during prolonged A-scope operation. The second study is by Mackworth (**37**) in England. His subjects were set to watch the movements of a pointer, a task thought similar to radar operation in some respects. Subjects missed seeing (or reporting) unannounced movements more and more as length of watch period increased. For an interesting interpretation of this kind of "fatigue," based on concepts of conditioning, the reader is referred to the original article.

Perceptual Problems

Discrimination of Separate Pips

Almost all of the preceding discussion has been devoted to the visibility of a single pip. In most of the studies, indeed, the pip was stationary. In actual operation, pips appear at different locations on successive antenna scans. If the scan is fast enough, the successive pip locations form a target course. They may cross the course of another target. As a pip closes in range it may also break up into two pips or three pips, that is, become evident as a cluster of separate ships or airplanes. At such times discriminability becomes important. Discriminability, of course, is basically a function of the resolving power of the CRT gun and screen. It may also be a function of the operator's visual acuity. With a given CRT and operator, pip discriminability would be a function of beamwidth and pulse length. The only published study directly relevant to this problem is that of Garner and Hamburger (**15**). They investigated pulse lengths from one to five microseconds in length on both a projection PPI and a standard PPI with a P7 screen, and found that above three microseconds of pulse length there was a minimum discriminable space which was proportional to the pulse length. Below three microseconds the P7 screen was slightly superior to the projection PPI. Their results are reported in terms of the minimum detectable range difference in yards between targets—a very practical unit. Extrapolating their data, it would appear that ultimate discriminability with short pulse lengths would be greater for the P7 screen than for the projection screen.

Visual Illusions

Special mention should be made of illusions, partly because of their inherent interest and partly because they tend to be ignored by non-psychological scientists. Illusions are real perceptual events caused by physical stimuli but they differ from other perceptions in their likelihood of misinterpretation. Two illusions have thus far been discovered in radar work.

The first, described by Sweet and Bartlett (**49**), is a subjective "second" sweep line on a PPI. It appears "real" enough, even though it is not as bright as the electronic sweep and assumes a different center of rotation, depending on the point of visual fixation. The "ghost" sweep has appeared thus far only at moderate-to-dim background brightnesses and at rotation rates above 15 to 20 rpm, although very high rates have not been investigated. It is believed to be a special case of the after-image known as Bidwell's ghost. It probably does not distort pip detection but does serve to distract the observer. The second is the linear-perspective illusion on the scopes used in GCA (Ground Controlled Approach) radar which will be referred to in Chapter 7. It is to be expected that more examples of illusions will crop up from time to time as radar undergoes technical development. Some may prove genuinely bothersome to the operator.

"Pipology"

In evaluating pips as to their probable target source, an operator uses cues of brightness, size, number, and location and especially the temporal variations in these. Add to these the operator's intelligence and experience and the result is "*pipology*," the fine art of target evaluation. In a sense, the whole knowledge of the radar system is part of it. The operator's available cues are ably discussd from the practical standpoint by Orman (**42**). Orman uses terminology familiar to radar operators and tells what to do and when to do it, using scopes current in 1947. Included in his article are hints on estimating the number and composition of sea and air targets from (1) size, (2) shape, (3) fluctuation and (4) movement of pips.

In another article, Orman (**43**) discusses the training of radar operators and strongly emphasizes the contribution of operator skill to the overall efficiency of the radar system. Operators really become trained in pipology, of course, and a fair amount of psychological research has already been expended on the problems of training methods. Training will no doubt change with the years, particularly with respect to training models and devices. For research on training methods carried out during the war, the reader is referred to the reports of Project SC-70, NS-146 of the Applied Psychology Panel, NDRC (**32**), and to Report No. 12 of the Army Air Forces, Aviation Psychology Program (**9**). Both references also deal with tests for selecting the best operators.

Summary

The general design of cathode-ray tubes is discussed in relation to the visual requirements of a human observer, with the chief emphasis on intensity modulated scopes of the type used in search radar. Radar scopes are rather poor as viewing screens, because they flicker and fluctuate in such a way as to make them difficult to maintain at optimal brightness. Even when they are operated optimally, they fall short of ideal visual requirements. Improvements to date have been notable but they have not been sufficient for two reasons: (1) there is a genuine lack of properly coordinated knowledge of the visual requirements themselves, and (2) the luminous output of electronic excitation of phosphor screens is not accurately predictable. Nevertheless, workers in both visual science and radar engineering have recently made significant advances as a result of direct visibility tests of various scopes. A major line of research is the investigation of the radar parameters which significantly determine visibility. Most of the research work in this field has been reviewed with the aim of highlighting the major problems and disclosing areas of ignorance. The major variables in visual detection are: the brightness and uniformity of the viewing screen; the contrast, shape and size of the pip to be seen; and the methods of visual search used by the observer. Many of these depend on the electrical parameters of the radar pulsing and receiving system. The functions relating the visual to the electrical variables are only partly known today, but those that are known are briefly described.

References

1. Ashby, R. M., Josephson, V. & Sydoriak, S. *Signal threshold studies.* Naval Research Laboratory Rep. No. R-3007, 1 Dec. 1946. Pp. 38.
2. Bachman, C. H. Image contrast in television. *General Electric Rev.*, 1945, **48**, 13–19.
3. Bartlett, N. R. The relation between critical flicker frequency and flash duration. *Amer. Psychologist*, 1947, **2**, 295. (Abstract)
4. Bartlett, N. R., & Williams, S. B. *Signal mark size and visibility of radar signals on a plan position indicator.* Systems Research Laboratory, The Johns Hopkins University. Rep. No. 166-I-30. 20 Sept. 1947. Pp. 16.
5. Bartlett, N. R., Williams, S. B., & Hanes, R. M. *Visibility on cathode-ray tube screens: The effect of size and shape of pip.* Systems Research Laboratory, The Johns Hopkins University. (In preparation)
6. Blackwell, H. R. Contrast thresholds of the human eye. *J. opt. Soc. Amer.*, 1946, **36**, 624–643.
7. Bramley, A. Aluminum backed phosphor screen in cathode ray tubes. *Electrochem. Soc.*, preprint No. 91-30 for meeting Apr. 9–12, 1947, 463–466.
8. Brill, E. R. *The improvement in minimum detectable signal in noise through the use of the long after-glow CR tube, and through photographic integration.* Harvard University Radio Res. Lab. Rep. No. 411-84, 8 Feb. 1944. Publ. Bd. No. 14158, U. S. Dept. of Commerce. Pp. 27.
9. Cook, S. W. (Ed.) *Psychological research on radar observer training.* AAF Aviation Psychology Program Reports. U. S. Gov. Printing Office, No. 12. 1947. Pp. 331.
10. Craik, K. J. W., & Macpherson, S. J. *The effect of certain operating conditions on the visibility of P.P.I. radar echoes.* 1945. Med. Res. Council, APRU, Psychol. Lab., Downing St., Cambridge. APU 16.
11. Crouch, C. L. The relation between illumination and vision. *Illum. Engng.*, 1945, **40**, 747–784.
12. Epstein, D. W., & Pensak, L. Improved

cathode-ray tubes with metal-backed luminescent screens. *RCA Rev.*, 1946, **7**, 5–11.

13. Ferree, C. E., & Rand, G. Further studies of the effect of composition of light on important ocular functions. *Trans. Illum. Engng. Soc.*, 1924, **19**, 424–447.
14. Garlick, G. F. J., Henderson, S. T., & Puleston, R. Cathode-ray-tube screens for radar. *J. Inst. elec. Engng.*, 1946, **93**, (IIIA No. 5), 815–821.
15. Garner, W. R., & Hamburger, F., Jr. Detectability and discriminability on a remote projection plan-position indicator. *Proc. I. R. E.*, 1947, **35**, 1220–1225.
16. Haeff, A. V. Minimum detectable radar signal and its dependence on parameters of radar systems. *Proc. I. R. E.*, 1946, **34**, 857–861.
17. Hamburger, F., Jr., & King, E. J. *A recording photometer and its use in studies of cathode ray screen displays.* Systems Research Laboratory, The Johns Hopkins University. Rep. No. 166-I-45, Jan. 1948. Pp. 19.
18. Hanes, R. M. *A scale of subjective brightness.* Systems Research Laboratory, The Johns Hopkins University. *J. exp. Psychol.* (In press)
19. Hanes, R. M., & Williams, S. B. Visibility on cathode-ray tube screens: The effects of light adaptation. *J. opt. Soc. Amer.*, 1948, **38**, 363–377.
20. Hardy, A. E. A combination phosphorometer and spectroradiometer for luminescent materials. *Electrochem. Soc.*, preprint No. 91-8 for meeting Apr. 9–12, 1947. 127–146.
21. Harriman, M. *Visibility on cathode-ray tube screens: Search time as a function of signal strength.* Systems Research Laboratory, The Johns Hopkins University. (In preparation)
22. Harriman, M., & Williams, S. B. *Visibility on cathode-ray tube screens: Positive versus negative signals on an intensity-modulated scope.* Systems Research Laboratory, The Johns Hopkins University. (In preparation)
23. Hopkinson, R. G. An examination of cathode-ray tube screen characteristics. *J. Inst. elec. Engng.*, 1946, **93**, (IIIA No. 5), 779–794.
24. Hopkinson, R. G. Visibility of cathode-ray tube traces in radar displays. *J. Inst. elec. Engng.*, 1946, **93**, (IIIA No. 5), 795–807.
25. Innes, R. B. *Method of measurement and some performance characteristics of P-14 screens with a note on manufacturers' specifications for tubes containing P-14 screens.* Prog. Rep. No. 5 (NDRC Rep. No. 14-586), 27 Nov. 1945. Publ. Bd. No. 25732, U. S. Dept. of Commerce. Pp. 13.
26. Johnson, R. E., & Hardy, A. E. Performance characteristics of long persistence CRT screens; their measurement and control. *RCA Rev.*, 1947, **8**, 660–681.
27. Lamar, E. S., Hecht, S., Shlaer, S., & Hendley, C. D. Size, shape, and contrast in detection of targets by daylight vision. I. Data and analytical description. *J. opt. Soc. Amer.*, 1947, **37**, 531–545.
28. Law, R. R. Contrast in kinescopes. *Proc. I. R. E.*, 1939, **27**, 511–524.
29. Leverenz, H. W. Phosphors versus the periodic system of the elements. *Proc. I. R. E.*, 1944, **32**, 256–263.
30. Leverenz, H. W. Luminescence and tenebrescence as applied in radar. *RCA Rev.*, 1946, **7**, 199–239.
31. Leverenz, H. W., *et al. Final report on research and development leading to new and improved radar indicators.* RCA Laboratories, 30 June 1945. Publ. Bd. No. 25481, U. S. Dept. of Commerce. Pp. 165.
32. Lindsley, D. B. *Final report in summary of work on selection and training of radar operators.* Res. Rep. No. 19, OSRD 5766, 24 Sept. 1945. Publ. Bd. No. 12165, U. S. Dept. of Commerce. Pp. 27.
33. Lindsley, D. B., *et al. Use of Philco trainer in training of A-scan oscilloscope operators.* Res. Rep. No. 3, OSRD 2096, 15 Oct. 1943. Publ. Bd. No. 18363, U. S. Dept. of Commerce. Pp. 27.
34. Lindsley, D. B., *et al. Effect of extraneous illumination on detection of signals on an A-scan oscilloscope.* Res. Rep. No. 13, OSRD 4264, 18 Oct. 1944. Publ. Bd. No. 18360, U. S. Dept. of Commerce. Pp. 14.
35. Lindsley, D. B., *et al. Radar operator "fatigue": The effect of length and repetition of operating period on efficiency of performance.* Res. Rep. No. 6, OSRD 3334, 4 Jan. 1944. Publ. Bd. No. 18366, U. S. Dept. of Commerce. Pp. 38.
36. Lindsley, D. B., *et al. Use of PPI flash reading trainer in training of Navy search radar operators.* Res. Rep. No. 17, OSRD 4831, 20 Mar. 1945. Publ. Bd. No. 18336, U. S. Dept. of Commerce. Pp. 12.
37. Mackworth, N. H. The breakdown of vigilance during prolonged visual search. *Quarterly J. Psychol.*, 1948, **1**, (I), 6–21.
38. Nottingham, W. B. *Excitation and decay of luminescence under electron bombardment.* MIT Rad. Lab. Rep. Sec. 6-45, 22 Jan. 1942.

Publ. Bd. No. 8070, U. S. Dept. of Commerce. Pp. 68.

39. Nottingham, W. B. *Proposed performance specifications for the P7 long-persistence cascade screen.* MIT Rad. Lab. Rep. No. 69-2, 12 Aug. 1942. Publ. Bd. No. 3932, U. S. Dept. of Commerce. Pp. 9.

40. Nottingham, W. B. *Conference on P7 cathode ray tubes held April 5 and 6, 1943.* MIT Rad. Lab. Rep. No. 62-7, 14 May 1943. Publ. Bd. No. 3967, U. S. Dept. of Commerce. Pp. 68.

41. Nottingham, W. B. *Comparison of P-7 screen test methods.* MIT Rad. Lab. Rep. No. S9, 14 Mar. 1944. Publ. Bd. No. 32744, U. S. Dept. of Commerce. Pp. 26.

42. Orman, L. M. Pipology. *Coast Artillery J.*, 1947, **90**, 31–36.

43. Orman, L. M. Training of radar operators. *Coast Artillery J.*, 1947, **90**, 22–27.

44. Payne-Scott, Ruby. The visibility of small echoes on radar PPI displays. *Proc. I. R. E.* 1948, **36**, 180–196.

45. Ridenour, L. *Radar system engineering.* New York: McGraw-Hill Book Co., Inc., 1947. Pp. 748.

46. Sutro, P. J. *The theoretical effect of integration on the visibility of weak signals through noise.* Harvard University Radio Res. Lab. Rep. No. 411-77, 4 Feb. 1944. Publ. Bd. No. 14151, U. S. Dept. of Commerce. Pp. 53.

47. Swartz, R. O. *Radar-beam simulator for training devices and minimum signal detectability study.* Tech. Info. Pilot (monthly edition covering only reports from contractors of ONR), No. 285, D-8. NRL S67-5. 1210-221/46. 16 Oct. 1946. (Abstract)

48. Sweet, A. L., & Bartlett, N. R. Visibility on cathode-ray tube screens: Signals on a P7 screen seen at different intervals after excitation. *J. opt. Soc. Amer.*, 1948, **38**, 329–337.

49. Sweet, A. L., & Bartlett, N. R. An illusory rotating sweep in viewing certain types of cathode-ray tube screens. *Amer. J. Psychol.*, 1948, **61**, 400–404.

50. VanVleck, J. H., & Middleton, D. *Theory of the visual vs. aural or meter reception of radar signals in the presence of noise.* Harvard University Radio Res. Lab. Rep. No. 411-86, 5 May 1944. Publ. Bd. No. 14159, U. S. Dept. of Commerce. Pp. 101.

51. Williams, S. B. The effect of certain illumination variables on the visibility of signals on cathode ray screens. *Amer. Psychologist*, 1947, **2**, 347. (Abstract)

52. Williams, S. B. *Visibility on cathode-ray tube screens: Screen uniformity.* Systems Research Laboratory, The Johns Hopkins University. (In preparation)

53. Williams, S. B. *Visibility on cathode-ray tube screens: Viewing angle.* Systems Research Laboratory, The Johns Hopkins University. (In preparation)

54. Williams, S. B., & Bartlett, N. R. Visibility on cathode-ray tube screens: Problems and methods. *J. Psychol.*, 1948, **25**, 401–417.

55. Williams, S. B., Bartlett, N. R., & King, E. Visibility on cathode-ray tube screens: Screen brightness. *J. Psychol.*, 1948, **25**, 455–466.

56. Williams, S. B., & Hanes, R. M. *Visibility on cathode-ray tube screens: Intensity and color of ambient illumination.* Systems Research Laboratory, The Johns Hopkins University. (In preparation)

57. Williams, S. B., & Harriman, M. *Visibility on cathode-ray tube screens: Area of search.* Systems Research Laboratory, The Johns Hopkins University. (In preparation)

58. *Report on influence of color contrast on visual acuity.* Eastman-Kodak Co., OSRD 4541, 1 Nov. 1944. Publ. Bd. No. 33247, U. S. Dept. of Commerce. Pp. 120.

59. *The comparison between signal and noise.* MIT Rad. Lab. Rep. No. 43-21, 29 Jan. 1943. Publ. Bd. No. 2851, U. S. Dept. of Commerce. Pp. 10.

PART II.

DESIGN AND ARRANGEMENT OF OPERATING EQUIPMENT

Chapter 5

THE DESIGN OF CONTROLS

D. R. CRAIG and D. G. ELLSON

Indiana University

Introduction

This chapter is intended as a general treatment of the relationships between control design variables and the operating characteristics of man-machine combinations. It is not the purpose of this chapter to present a complete catalogue of information concerning the known effects of manipulating the numerous variables of control design. Much of this information is available in an annotated bibliography by Macmillan and Lawrence (**18**), in an extensive review by Jenkins (**17**), and in a less exhaustive treatment by Channell (**4**). These reviews may be supplemented by a very practical check list for the evaluation of equipment design (**21**) in those cases where immediate application of current information is desired. In addition, much of the material in the next two chapters applies directly or indirectly to the design of controls. Rather, it is the purpose of this chapter to present a systematic approach to the problems in this area in the hope that it will lead eventually to the basic research which is necessary to relate and reconcile many of the facts brought forth by the recently accelerated interest in equipment design.

Historical Approaches to the Problems of Control Design

The literature pertinent to the design of manual (and pedal) controls is not very extensive, but the heterogeneity of the subject-matter and the specificity of most of the research make summarization difficult and of questionable value. The implied definition of "controls" has embraced any and all devices manipulated by the operator of a machine or instrument. It is obvious that the problems involved in the design of the knob on a range-selector switch have little in common with the problems involved in the design of a remote control gunsight. Both devices, however, qualify as "controls" and usually enter into discussions of control design on fairly even terms. Likewise, there has been a failure to distinguish between design problems related to the nature of the operator's total task and design problems related to the operating characteristics of the control device. As a result, in discussions of control design one finds such factors as the color coding of control knobs grouped together with such factors as gear ratios and inertia. The types of problems and classes of variables considered by any one group of investigators have been determined simply by the interests and responsibilities of that group, rather than by the demands of a systematic approach.

During and immediately following World War II, a considerable amount of research was performed on the effects of the characteristics of the task on control operations involving a human operator. It was stimulated largely by requests for data from engineers and designers of military, aviation, and other types of complex equipment which require highly skilled performance for efficient and accurate operation. It included the determination of the effects on speed and accuracy of control operations of such variables as size, shape, and location of control knobs; friction, damping, and inertia of controls; operating radii of controls; and gear ratios. Much of this work was necessarily exploratory and most of that which was performed during wartime suffered from restrictions on experimental design imposed by the use of modified service equipment and by the need to answer specific questions about specific types of equipment.

Other wartime research emphasized the modification of existing equipment in the direction of ease of learning, economy of effort, and accuracy in using the right control at the right time. This research was mainly stimulated by those responsible for the safety and training of personnel.

In the years preceding World War II a related type of research on design and placement of controls had been carried on by industrial psychologists and time and motion engineers. Their major interest, however, was in the evaluation of motor activity for specific work positions in order to eliminate waste motion and in the alteration of performance in the direction of increased production.

It is well recognized that the approach that is adopted for any broad research program greatly affects the nature of the results which are obtained. In the general lines of research just discussed the approach was essentially that of attempting to improve the performance of certain existing man-machine combinations by modification of the equipment in current use. This approach is a natural outgrowth of the history of machine design.

Traditionally, the behavior of the human operator was considered unpredictable and outside the realm of description in engineering terms. On the other hand, considerable confidence was retained in the ability of the operator to adapt to the requirements of the operation. It became standard practice to ignore operator characteristics in the design of machines. The machines and their controlling mechanisms were designed without specific reference to the operator and then these systems were linked to a control device which, if manipulated in a certain prescribed manner, would provide the desired input to the rest of the system. Sometimes the assumption that the average operator could and would manipulate the control device in the prescribed manner was justified. More often, however, it was found that the output of the man-machine combination was somewhat less than that which was desired.

The problems posed by the human operator were complicated by the apparent necessity for using him to perform certain functions. One approach to a solution was the adoption of a general policy of reducing the role of the operator to a minimum. The rapid development of the engineering field led to more and more success in carrying out this policy. However, even in his "minimum" role the operator continued to be a problem.

Meanwhile, the observation of differences in the abilities of operators and the effects of practice stimulated another general attempt to deal with the problems of the human operator. It was found that the output of the man-machine combination could be improved and made more consistent by means of selection and training of operators. This attempt to "fit the man to the machine" was recognized as an incomplete solution, but the results were gratifying enough to warrant continuance and expansion of the program.

Another method of dealing with the same general problem was suggested by observation of the difficulties which the operators experienced with a specific piece of equipment. It was noted that by making certain obviously desirable changes in the equipment the performance could be improved. The success of this procedure based on casual observation and trial-and-error tactics led to more detailed analyses of the task, more carefully controlled observation, and the application of statistical checks on the validity and reliability of the results obtained from the changes which were made. The investigations noted at the beginning of this section are, for the most part, the result of this method of approach.

Recently, technological developments stimulated by the demands of World War II have highlighted the inadequacy of these approaches to the problem. The complexity of much of the equipment has required operator skills which can hardly be termed a "minimum role." This same trend toward complexity has resulted in the design of equipment which demands manipulations that even carefully selected and highly trained personnel cannot perform satisfactorily. These developments, along with the expense in time and money involved in elaborate programs of selection and training, have emphasized the need for including operator characteristics in the original design layout. Solution of the general problems of equipment design based on this approach will depend upon the establishment of broad research programs founded on a consideration of (1) fundamental operator-control relationships and (2) the relationships between operator-control combinations and the over-all task.

The Operator as an Element in a Control System

In designing a control system for a specific application, the designer necessarily has a clear, definite goal in mind; the system is to perform some given task, and it must do so with some minimum desired quality of performance. Although the control system in some instances may consist of only one simple element, more often it consists of a number of elements linked together. The designer today has a choice of innumerable devices from which to draw elements for his system. His choice depends to a great extent on the nature of the task which the system must perform. For example, if rapid control action is required, certain classes of elements and linkages are immediately rejected as too slow. If the maximum load is large, certain low-power or fragile elements must be rejected or, if they are included, other elements must be incorporated into the system in a way which will prevent the low-power elements from being subjected to extreme loads. In addition, the designer's choice will be affected by the conditions under which the system will be required to operate. Extremes of temperature, probability of jolting, space restrictions, etc., are examples of such conditions. In general, the designer needs to know the nature of the task and the expected operating conditions, on the one hand, and on the other he needs to know the operating characteristics of the available elements, the operating limits of these elements, and the susceptibility of these elements to the expected conditions.

The control systems in which we are interested in this chapter are those systems which include a human operator as one of the elements. In line with the approach here adopted, the operator is an element to be considered in the original design of the system. If the designer is to include the operator in the design, he will require the same sort of information about the operator as he does about the other elements. He will need to know his operating characteristics and limits and his susceptibility to task conditions.

Certain features of the human operator as a system element need to be pointed out. First, the inclusion of the human operator is not usually a matter of choice but rather a matter of necessity. As an element he must be accepted with all his limitations and characteristics, and the choice of other elements becomes largely a matter of compensating for and supplementing these characteristics and limitations. Second, differences between operators are greater than is the case with other classes of elements. Third, unlike other elements, frequent interchange of operators is the rule. Fourth, the operator is seldom used simply as an element in a single control system. Ordinarily he functions intermittently as an element in several systems and, in addition, may be required to perform other functions than those of control. Fifth, the link between the operator and the rest of the system often is not fixed or controlled as are other links in the system. This arises mostly from the practice of including many unit tasks in the over-all task of a single operator. These are all features which the designer must consider in the case of the human element but which rarely have their counterpart in the case of the other elements. They should be viewed as unique features, however, rather than as insurmountable difficulties.

The relationship between these features and the problems posed by operating conditions also deserve comment. To some extent, the designer of a control system may counteract the disadvantages inherent in operating conditions by selecting equivalent elements which are not adversely affected by those conditions. In general, this principle can not be used in the case of the human operator, for, as was stated above, he is usually included as a matter of necessity rather than choice. The designer may also build protective devices into his system to eliminate or diminish the effects of adverse conditions. This principle is applicable to the human element as well as to the others. However, the last three "unique features" listed above are essentially operating conditions, and the manner in which they must be taken into account is not covered by either of the above principles.

The effects of these conditions which are inherent in the use of a human element may be modified by fitting the manipulative requirements to the expected conditions. For example, in many military tasks the operator is under stress and his characteristics may tend to change as a result. The degree of this stress varies. If a gunner is tracking an approaching plane, the nearness of the plane may increase the stress and initiate certain reflex or habitual responses. Adjustment of the control device should be arranged to coincide with these stress reactions rather than to oppose them. Again, the fact that the operator can not be "bolted down" dictates against the requirement of protruding devices, if sudden displacement of the operator is a probability. Also, such principles as standardization of controls, shape coding, and design of movement sequences in the over-all task may influence the choice between a rotating knob and a lever device. In short, although the designer's primary concern is the accomplishment of the proper arrangement of elements such that control data which are fed into the system at one end will be transformed into the proper output at the other end, he must also be concerned with the expected operating conditions when selecting the elements for his system.

In the development of equipment requiring a human operator, the problems connected with the design of a specific control system which forms a part of the equipment may have their source either in the nature of the control task or in the expected operating conditions. These two factors are interrelated and the importance of the problems relative to each varies from case to case. The next section will include a discussion of these interrelationships, a classification of problems according to source, and a classification of control tasks.

The Relationships of Design Problems to Design Criteria and to Types of Control Adjustments

Design Criteria and the Classification of Design Problems

The great variety of tasks in which the operator manipulates "controls" and the great variety of control devices combine to produce innumerable design problems. For purposes of discussion and systematic research it is necessary that these problems be classified. Before examining a suggested classification let us consider an hypothetical case of design evaluation.

The placement of the control in relation to the operator is included in this discussion as one of the variables in control design. As such a variable its evaluation may serve to illustrate some of the relationships between the problems which arise in the design of controls. Specifically, we will consider the placement of a handwheel in respect to the operator for a task in which manipulation of the handwheel is an intermittent operation. It will be assumed that the handwheel is the proper device for the required adjustment. In the process of evaluating a tentative position of the handwheel, three pertinent questions may be asked.

1. Does the location of the handwheel avoid waste motion or uncomfortable positions for its operation?

2. Is the handwheel placed so that the operator can grasp it quickly and accurately without excessive attention to its exact location?

3. Does the position of the handwheel allow for the optimal range and type of movement which the adjustment requires?

In attempting to answer these questions several important implications become evident. It is obvious that in many cases the answer could not be an unqualified "yes" to all three questions. Yet, in so far as they are all three pertinent questions they would all need to be considered and some compromise effected.

In addition, it should be noted that each of the three questions is based on a different set of criteria for evaluating a single design variable. The criteria giving rise to question (1) emphasize the effects of the design on the over-all efficiency and welfare of the operator. Question (2) is stimulated by a set of criteria which emphasize the speed and accuracy with which the operator may contact his controls. Question (3) is based on criteria which emphasize the effects of the design on the adequacy with which the control may be manipulated relative to the adjustment required.

Again, it should be noted that the relative weight given to the three sets of criteria will depend upon the relation of the control operation to the expected operating conditions and upon the type of adjustment for which the control is used. If the control is operated very infrequently and for only brief periods, then the answer to the first question is not very critical. If the control normally is operated at a time when other demands of the task are minimal and is not involved in necessary adjustments to "emergency conditions," then the answer to the second question is relatively unimportant. Finally, if the necessary adjustment of the control is of the "on-off" type requiring a single initiating movement, then the answer to the third question is greatly reduced in significance.

TABLE I

CLASSIFICATION OF EQUIPMENT DESIGN PROBLEMS

Class No.	Type of problems	General factors to be considered
I	Problems relative to maintenance of over-all operator efficiency	Ease of learning; economy of effort; facilitation of comfortable, safe operations; etc.
II	Problems relative to the speed and accuracy of contacting controls	Ease of tactual or visual identification of controls; accessibility of essential, frequently used controls; degree of interference between successive or coincident operations; etc.
III	Problems relative to control manipulation as an element in a control process	Suitability of transmission properties of the control system

Also, it should be pointed out that the considerations which prompt the first two questions are based upon evaluation of the design variable in respect to its relationship to various operating conditions, whereas the third question is concerned solely with the effect of the variable on the required adjustment.

The problems cited in the above example

were distinguished on the basis of different sets of criteria for evaluating the effects of design variables on the operator's performance. That principle has been used as the basis for a general classification of design problems as presented in Table I.

The corollaries of this classification include generalized statements of the implications noted in the handwheel placement example.

1. A single design variable may be involved in problems of more than one class.

2. The optimal values prescribed for a given design variable by different sets of criteria may not always coincide.

3. In any particular application the importance of a given class of problems varies according to the relation of the control operation to the expected operating conditions and according to the type of adjustment which is required.

4. Problems of Classes I and II arise from a consideration of effects of operating conditions. Problems of Class III arise from a consideration of operator movement characteristics and the requirements of the control task.

Types of Control Adjustments and their Relation to the Classes of Design Problems

The nature of the control task constitutes the starting point for the design of a control system. In the paragraphs above it was noted also that the nature of the adjustment (or control task) partially determines the relative importance of the different classes of design problems which may be met in any specific instance. Table II presents a classification of control adjustments, along with an additional column indicating the classes of design problems which usually arise in connection with each type of adjustment.

Certain devices are typically used to accomplish certain kinds of adjustment. In the following discussion this fact will be utilized to illustrate the type of adjustment being considered. However, it is important to note that it is types of control adjustment that we have classified, and not types of devices manipulated by the operator. For example, a toggle switch may be the control by which power for a drill press is turned on or off. The same toggle switch may be used to turn on or off the power to a motor which positions a gun turret. In the former case the switch is used for an on-off adjustment (Type A-1). In the latter case the switch is used for a finite setting adjustment (Type B).

Adjustments of Type A-1 normally engage or disengage a source of power and thus control its availability for further operations. The ignition key and starter button of an automobile are examples of devices requiring this type of adjustment. Various clutch arrangements in machine shops and automotive vehicles are further examples, as are light switches and the power switches on innumerable electrical appliances.

TABLE II

Classification of Control Adjustments

Type	Adjustment	Class of problems usually encountered
A-1	On-off	I, II
A-2	Finite setting–aided	I, II, (III)
B	Finite setting–unaided	I, II, III
C	Continous correction	I, III

Type A-2 adjustments involve the selection of different amounts of power or different operating equipment. As with Type A-1, the adjustment consists of effecting a desired setting of the control device, but in the present case a greater choice of settings is available. Wrong settings may more easily occur and sequences of settings may be required. In both types the precision of the adjustment is accomplished by mechanical features of the equipment and thus successive adjustments are not normally a function of the precision of preceding adjustments. The levers controlling the speed of submarine engines are examples of devices used for power selection. Examples of equipment selection devices include gear-shift levers, knobs for selection of different radio frequency bands, and knobs for changing the operating range of radar equipment.

Before continuing to the other categories it might be well to remind the reader that all classifications are subject to the difficulties posed by the "borderline case" and that this classification is no exception. However, these difficulties do not impair the usefulness of classifications for purposes of general discussion.

The adjustments classified as Type B are finite setting adjustments in which the precision of the setting is dependent upon the operator. Continuous gradation of position is provided and thus precision depends on the operator's manipulations rather than on mechanical positioning aids. Normally, the frequency of settings is such that secondary corrections may be utilized to compensate for errors in the initial positioning attempt. The data for these secondary corrections arise from the difference between the desired setting and the actual setting initially accomplished. However, the data for the primary settings arise from other sources and are not a function of the preceding adjustment. Typical examples of this type are the tuning adjustments in the operation of a radio, the focusing of optical instruments on fixed objects or for fixed ranges, and the operation of some models of torpedo data computers. The toggle switch example at the beginning of this discussion illustrates the use of remote control and power amplification to accomplish this type of adjustment.

Adjustments of Type C are continuous correction adjustments such as tracking or steering operations. These adjustments are error-reducing responses. The data are presented to the operator as a misalignment between two points and the adjustment consists of reducing the misalignment to zero. Changes in the extent of the misalignment are only partially under the control of the operator. This is one of the principal characteristics which distinguish Type C from Type B adjustments. Type B adjustments are movements of a variable point to a fixed position and the movement of the point is completely under the control of the system through which the adjustment is made. In Type C adjustments, however, one or the other of these conditions does not obtain. The position to which the variable point must be moved may be continuously changing, as in pursuit tracking, or the variable point, although linked to the operator, may be subject to forces external to the control system, as in compensatory tracking. The differences between tracking systems will be analyzed further in the later sections.

The relationships between types of adjustments and classes of design problems as indicated in Table II require little comment. Problems of Class I are encountered in all types of adjustments, just as they are encountered in all human tasks. Problems of Class II are usually encountered in all types of adjustments because of the current practice of including several operations in the task of any one operator. The exception is Type C in which the need for continuous adjustment usually restricts the task to a single operation or requires that additional operations be accomplished without interruption of the operation involving Type C adjustments. The problems of Class III, however, bear a somewhat less general relationship to the several types of adjustments.

The extent to which Class III problems are involved in a specific control adjustment is proportional to the sensitivity of the control system to operator movement characteristics. Most control systems requiring adjustments of Types A-1 and A-2 are relatively insensitive to the movement characteristics of the operator. The rate at which successive adjustments are required is low and the deviations of variables such as the extent, force, speed and direction of the movement are either controlled or filtered out. The reverse is true at the other end of the classification. Type C adjustments usually are accomplished through control systems which are extremely sensitive to the operator's movement characteristics. It is in this area that further research is most sorely needed. It is also in this area that the need for a consideration of operator characteristics has been demonstrated most critically to those who are responsible for the development of equipment intended for human operation.

An Evaluation of Current Information on Problems Relative to the Design of Controls for Human Use

The shift of emphasis from ignoring the operator in equipment design to the inclusion of operator characteristics and operating conditions in the original design layout has resulted in a need for a great deal of information about the operator. As was stated earlier, the designer needs to know the movement and transmission characteristics of the operator, he needs to know his operating limits, and he needs to know his susceptibility to expected operating conditions. As was also stated earlier, the designer needs to be aware of the principles by which the operating conditions arising from the over-all task design may be modified.

The detailed treatment of operating conditions is not strictly within the scope of the present chapter. The chapters on panel layout and arrangement of equipment are concerned with the principles of over-all task design. Other chapters deal with the effects on the operator's performance relative to environmental factors, stress, fatigue, etc. Problems arising from these sources were classified as Class I and Class II. They have been included in the discussion thus far partly to provide perspective for the problems of Class III and partly because a control system originally designed or later modified in isolation is very likely to possess features which are unsatisfactory in terms of actual operating conditions. The design of a control system may increase the difficulty of the over-all task as well as the difficulty of the control operation. Increased difficulty complicates the problems of selection and training. The design of a control system may demand unnecessary effort on the part of the operator. Even though this added increment of effort is not in itself significant, it becomes significant when considered in terms of the total demands made on the operator by the over-all task. The design of a control system may also demand more of the operator's attention and abilities than is necessary. Such demands may restrict unfavorably the scope of the over-all task which may be assigned to the operator. It is for these reasons that the problems related to operating conditions have been included in the preceding discussion. However, the rest of the chapter will be concerned primarily with the problems of Class III.

The restriction of the discussion to problems of Class III also operates as a restriction on the types of adjustment to be considered. For adjustments of Types A-1 and A-2 a consideration of the manner in which the demands of the system fit into the operating conditions far outweighs the influence of operator characteristics on the adequacy of the adjustment. We will therefore be dealing, for the most part, only with adjustments of Type B and Type C.

The problems of Class III relate to the suitability of the transmission characteristics of the control system. In respect to the human element in the system, the response which the operator makes to the control data which he receives must result in a movement of the control device that is adequate for the accomplishment of the required adjustment. In general, the operator is the potential source of innumerable movement patterns. He is also capable of receiving control data in many forms. What are the transmission possibilities of the operator? What are his transmission characteristics? How best may his transmission possibilities and characteristics be exploited for a given type of control task? Unfortunately the answers to these questions are not readily available.

The practice of studying the effects of modifications in existing equipment has directed attention away from the basic question of whether or not the control system being used is the best one for the task. For illustration we may consider the task of making settings on a linear scale. This is a Type B task. A study of optimal values for a rotating knob device indicates that by the use of an optimal knob diameter, an optimal ratio between control movement and pointer movement, and an optimal error tolerance, subjects were able on the average to complete an adjustment in from 2.0 to 2.5 seconds depending on the distance to be traveled to the required setting. However, studies of position adjustments using a sliding device consistently report total adjustment times under 1.0 second. Thus, in a Type B task in which minimum time for a total adjustment is desired, a sliding device for positioning the pointer will result in faster adjustments than an optimally designed rotating knob device. The results of the former study will furnish the designer with information concerning the speed of performance to be expected with a rotating knob for making linear scale settings and will furnish him with optimal values for various design variables involved in such a device. On the other hand, it will not tell him the optimal design factors for "making settings on a linear scale." Information of the former sort is typical of the information now available. Information of the latter sort is typical of the information which is lacking. Device manipulation rather than the adjustment of the control system has been considered "the task."

A second general feature of studies on control design is the failure to control other variables in the system while modifying the variable about which information is desired. Studies of compensatory tracking (a Type C adjustment) have reported optimal gear ratios in terms of required winding speeds when handwheel cranks are used. The optimal gear ratios are those which allow the fastest winding speeds over the range of target speeds expected. Tracking accuracy increases with winding speed, but the latter is limited by fatigue factors and the maximum speed attainable. In this task the operator receives most of his control data from the observation of the position and movement of a variable pointer relative to a fixed pointer. The smoothness of the movement of the variable pointer and the rate at which it moves affect the nature of the data which the operator receives. These two factors are also related to the gear ratio of the handwheel device. The question therefore must be asked as to whether the fast winding speeds are optimal because they accomplish optimal movement characteristics or whether they are optimal because they improve the display characteristics of the system. Fast winding speeds have their disadvantages. If they merely improve display characteristics, then they are optimal only insofar as no other means of improving the display is available. Further research is called for.

A third unfortunate feature of studies on control design variables lies in the interpretation of the results. Conclusions are stated in terms of operator performance, although the data are obtained from system performance. Consequently, the differences between control tasks as well as the differences between control systems tend to be neglected. The effects of neglecting these differences may be illustrated by the reported effects of adding friction to manipulative devices. In the literature the results of a number of different studies have been summarized as the effects of friction on operator performance and have been considered contradictory. However, the results are actually the effects of introducing or increasing friction at a certain point in different control systems. The situation is analogous to an attempt to state the effects of increasing the octane rating of the fuel on the performance of gasoline engines. If no account is taken of the differences between engines (systems), the results will be equivocal. On the other hand, if differences in the characteristics of the several types of engines are considered, then meaningful statements concerning the relationship between octane rating and engine performance become possible. There is nothing contradictory in the fact that the best octane rating for airplane engine fuel is different from the best rating for fuel used in the average passenger car.

The effects of manipulating a variable in one system are not indicative of the effects to be obtained with similar operations on another system unless (1) the two systems are quite similar in respect to their essential characteristics or (2) a known relationship obtains between the variable and the characteristics in which the two systems differ. Control systems may contain essentially the same physical elements and yet differ considerably relative to the control task and relative to the role of the operator in the system. The lack of specific reference to the operator in traditional designs has obscured the apparent need to analyze these differences in the study of design variables.

In addition to these criticisms of the general approach employed in most control design studies there are others which relate to deficiencies in the application of scientific procedure. These are: (1) data based on too few subjects, (2) data based on subjects with widely varying practice histories, (3) incomplete statistics, and (4) faulty experimental design. As a result the available data are limited in their usefulness and must certainly be applied with caution. At best they may be used as hypotheses for further research and as substitutes for blind hunches in cases where equipment modifications are contemplated. Their application should be considered with full knowledge of the conditions under which they were obtained. The results of their application should be tested by controlled experiments using mock-ups or working models, including as much of the total task as possible, and giving due attention to the problems of all three classes.

Many of the pressures and conditions which led to deficiencies in procedure in the earlier studies are no longer operative. Likewise, recent studies dealing with Class III problems have indicated the adoption of a new approach similar to that presented in this chapter. In general, the recognition of the designer's need to include the operator in the design of control systems has been accompanied by a recognition of the need for a different sort of information than is currently available. Attention is now directed toward system analysis and toward analysis of the operator's role in the different systems. A brief orientation to this point of view is presented in the following section.

System Variables and The Role of the Operator in Manually Operated Control Systems

Throughout the previous sections of this chapter we have considered the operator as one of the elements in a control system. In a very general fashion his role has been described as the transmission of control data to the rest of the system by means of the manipulative devices provided him. Actually the details of his role differ from system to system, and, conversely, the details of the operator's role may determine the classification of the system. It is necessary therefore, to analyze the ways in which his roles may differ before effective use can be made of his characteristics, before system differences may be accurately appraised, and before results of system modifications may be properly interpreted. Because of the relative novelty of the present approach, a few general statements concerning transmission systems prior to beginning the discussion of system variables may help to clarify the terminology.

Most transmission systems consist of a number of units linked together. Although each unit is a transmission system in itself, to avoid confusion we shall call these units "transmission elements" and reserve the term "transmission system" for a combination of these elements. Generally, in such combinations the elements are so linked that the output of one element constitutes the input to the next element in the series. Thus, in the present case the control data constitute the input to the operator and the resulting movement constitutes the output of the operator as well as the input to the control device. An additional type of linkage may occur in which the output of an element at some point in the series is fed back into the system at some previous point in the series. Systems containing such linkages are appropriately called feedback systems. In the discussion which follows we will have occasion to consider a special class of systems of this type, namely, servo systems.

"A servo system is a feedback system in which the actual output is compared with the input, which is the *desired* output, and the driving element is activated by the difference of these quantities" (**16**). There are numerous examples of man-machine combinations in which the operator controls the output of the machine by comparing the actual output with the desired output and then making a compensatory adjustment of the controls to correct the difference. The control of a plane's attitude or a ship's heading or a reticle's position in respect to the target are all familiar examples of closed-cycle control systems, or servo systems, in which the human operator serves as a biomechanical link. Although it is almost impossible to find a humanly operated control system in which the operator receives no information concerning the result of his manipulations, the manner and degree to which this feedback affects the operations may vary from the critical role played in the examples above to the minor role exemplified by the radio operator's casual check of his instrument to make sure that turning the power switch actually turned on the power.

The output of a transmission element depends on three factors: (1) the characteristics of the input, (2) the transmission characteristics of the element, and (3) the characteristics of the load against which the element operates. Thus, the output of the operator is determined by the nature of the control data, by his own transmission characteristics, and by the characteristics of the load inherent in the design of the device which he manipulates. System variables involved in each of these factors will be taken up separately, but the interrelationships which exist between them must necessarily be considered in any research work or practical application.

The concept of "the operator" also needs a brief comment. At the level of analysis employed thus far in this chapter the operator, as an individual, has been considered a transmission element. However, for design purposes it is important to realize that he is actually a complex network of numerous transmission systems. He possesses a number of different types of receptors or sense organs each of which is peculiarly sensitive to a certain class of inputs. Likewise, his fingers, toes, hands, feet, arms, legs, and head are all outlets for utilizable movement patterns. His possibilities are further multiplied by the fact that the correlation of a given input with a particular output (movement pattern) is not "built into" the operator except perhaps at the reflex level. It is only by means of instructions, training, and system design that certain of his movement patterns may be standardized for a given class of inputs in a particular situation. Recognition of the multiplicity of transmission systems embodied in the operator necessarily leads to the realization that input design and "operator" design deserve the same detailed consideration that has recently been focused on the design of manipulative devices, for the role of the operator is determined by the interrelationships between design characteristics of each of the three elements. This fact is important not only to the designers of control systems but also to the research worker who must evaluate the role of the operator in appraising the similarities and differences between systems.

In the following paragraphs some of the more important system variables are listed together with brief orientative comments.

Input Variables

Sense Modality. It was stated above that control data may be furnished to the operator in several forms because of the differential sensitivity of his sense organs. The sense modality through which the data are transmitted determines certain relationships relative to the operating conditions as well as to the control operations. Very often the selection of sense modality is not a matter of choice for the designer. However, it is sometimes possible to solve many problems caused by auditory data presented in a noisy situation by switching to visual presentation. It is also possible in some tasks to relieve the load on the operator's visual equipment by switching the presentation of data to another modality. The development of "Flybar" is an interesting example (**5**). In order to carry out such design manipulations it is necessary to know the characteristic differences which obtain between the several modes of input presentation.

Number of sense modalities employed. More than one sense modality usually is involved in a single control system. Tactual and kinaesthetic data in the form of feedback are often important accompaniments of the visual or auditory data received from the primary display. The feedback may be from the manipulative devices (the "feel" of the controls) or from the motion effects which the operation causes (the "seat of the pants" data to which the aircraft pilot refers). The possible provision of additional control information by tactual or kinaesthetic feedback should be considered in original system designs. The recognition of these feedback inputs is important also when modifications are considered which distort or exclude these data. For example, the introduction of booster systems between the pilot's controls and the control surfaces of the aircraft has led to a need to impart "artificial feel" to the controls (**19**). The contribution of these feedback data must also be considered in connection with experimental adjustment of inertia, damping, friction, etc., in manipulative devices. The "fatigue" effects of prolonged operations also may involve the distortion of these feedback data through "numbness" or the adaptation of the sense organs involved, although information on this point is lacking.

For convenience of exposition the comments concerning other input variables will be in reference to visual inputs only. However, these variables apply generally to inputs in all sense modalities.

Directness of data. This variable may best be described by example. In a Type B adjustment involving the linear positioning of a pointer the data may be furnished by displaying verbally or visually a value corresponding to a scale reading. By referring this data to the scale the operator "sets up" the new position which is required. The rest of the operation consists of aligning the pointer with the new position. The indirectness of this input system may be compared with a display of the data in a manner which "sets up" the new position directly. This may be done by signal lights behind or above the scale divisions. Or a movable marker such as a pointer or spot of light may be used. The differences between these display techniques may affect total response time as well as the number of sources of error in the total operation.

Continuity of data. The input for Type B adjustments is characteristically intermittent, but the input for Type C adjustments is generally intended to be continuous. However, physically continuous data may be effectively intermittent. If the velocity of pointer movement constitutes a portion of the input in a tracking adjustment, there are finite limits outside of which the velocity is not perceptible. The continuity of the velocity data as effective input will depend therefore on whether or not the operating range of pointer velocities remains within these limits. (See also "Operating range of input magnitude.")

Homogeneity of data. Many tracking displays vary according to the number of input characteristics which may function as effective data. In pursuit tracking[1] the target and reticle are both visible and thus the "error" or relative position is visible. The target exhibits position, velocity or rate, and acceleration data. So does the reticle. The "error" exhibits size, rate of change of size, and acceleration of size changes. The reticle may be instantaneously "on target" in position but off in rate, so that a positional error immediately develops. The reticle may be "on" in rate but off in position, so that a constant positional error exists as long as rate is constant. Other possibilities are obvious. In compensatory tracking only the error and its derivatives are visible. Displacement of a marker from a null position indicates error. This technique also allows separation of target movement and its derivatives. The "error" may be in position, rate, or acceleration. If it is rate indication, then the operator knows nothing about position except when it is constant but may receive acceleration data from the rate of error development. If it is acceleration indication, then he knows nothing about position and nothing about rate except when it is constant. The rate of error development now reflects higher derivatives of target movement which are probably negligible for purposes of control.

The homogeneity of the input may be regarded in three ways: (1) in terms of the amount of information displayed, (2) in terms of the degree of complexity in the operator's transmission task, and (3) in terms of the possibility of intermittent selection of effective input from the total input pattern.

Operating range of input magnitudes. The human operator is characterized by definite limitations concerning the perception of changes in a variable as well as the perception of the variable itself. We mentioned above that pointer velocities could exceed perceptual limits. We are here concerned with minimal values or the "thresholds" below which the perception of differences is impossible. In a compensatory tracking task the error indicator may move from the null position by so slight an amount that the positional change is not effective as an input. Or it may move at such a slow rate that the latter is not perceptible. Or it may move at a variable rate but with the variations (accelerations) not perceptible. The derivatives of target position to which these display variables correspond may be utilizable as effective inputs in direct tracking, but the display characteristics may have filtered them out by means of the operator's perceptual thresholds.

The continuity of the input as well as the homogeneity of the input may be affected by these perceptual thresholds. The average rate of a translatory displacement increases with the amplitude of the displacement. If the operating range of the amplitudes is great enough, the amplification or reduction characteristics of the transmission system between the operator's movement and the pointer which he moves may be adjusted so that (1) the rates of all pointer movements exceed perceptual limits, or (2) rate is perceptible over a portion of the amplitude range, or (3) the rates of all pointer movements are perceptible.

Variables in experimental input patterns. Certain variables in input patterns have been observed in experimental work which are worth mentioning, although their applica-

[1] In pursuit tracking, sometimes called direct tracking, the operator attempts to superimpose an indicator on a moving target which is visible to him. In compensatory tracking the operator seeks to keep an indicator from drifting away from a stationary zero position.

bility to practical situations is not yet known.

1. Range effect. It has long been known that judgments of the magnitude of a variable are distributed over a narrower range than the input magnitudes. That is, large inputs are underestimated, small inputs overestimated, mean inputs unaffected. This phenomenon has recently been observed in experimental investigations of responses to sudden displacements of the "target." In the Naval Research Laboratory experiment (**20**), the equipment was similar to a compensatory tracking system. A bright dot on an oscilloscope, centered on a hairline, was suddenly displaced and the operator was required to make a translatory movement with a lever type device to bring the dot back to the hairline as soon as possible. In the Indiana University study (**11**) the "dot" formed by a vertical line moving behind a narrow horizontal slit was suddenly dispaced and the operator was required to realign his sliding pointer with the target "dot" as quickly as possible. In both of these studies large displacements were undershot on the average and small displacements were overshot on the average. Results from another study at Indiana University (**12**) show a significant difference between the mean amplitude of response to a one-inch displacement when that displacement is the largest of a sequence of variable displacements and when it is the smallest of a sequence. The former two studies were designed to investigate the response to a step-function input in continuous tracking (Type C). However, it may be that these responses are more accurately described as Type B or "setting" adjustments.

2. Predictability. This variable of the total input sequence is still in the status of a reasonable hypothesis. The results of the Indiana University studies seem to indicate that regularities in the input sequence allow the operator to predict certain aspects of future inputs. Such prediction adversely affects the proportionality between the input and the output of the system. The importance of the relationship will be discussed later.

Operator Variables

We have noted above that the correlation between input and output in the case of the operator can not be "built in." We have also noted that "the operator" is an abstraction of the transmission functions of a human being in a particular situation. The degree to which the designer's abstraction fits the individual operator in any given instance is affected by the variability between operators and the variability within one operator as a function of time. A few brief statements are therefore pertinent before considering the system variables related to the operator.

Variability of operator characteristics in respect to a particular system design involves changes due to a change of operators, changes accompanying practice, and changes due to fatigue. By means of selection and training the range of individual differences may be reduced considerably. By designing the system in terms of the "average operator" the effects of these differences may be further reduced. The effects of practice may be controlled to a satisfactory extent by utilizing the characteristic form of the "learning curve." The development of standardized motor responses follows a decelerating course. By training operators to a point where little if any change results from additional practice and by maintaining them at that level, learning effects may be minimized. (It should be noted that this principle carries implications for the designer and the research worker relative to the differences between characteristics of trained and untrained operators.) Finally, the effects of fatigue may be reduced by control of the operating environment, by attention to the difficulty and duration of the task, and by the conditioning and adaptation which accompanies the regular performance of a task.

In control systems the outputs of the operator which are standardized for a given class in inputs consist of movement patterns. These patterns may vary in terms of duration, extent, rate and force. The effective operating range for each of these movement characteristics differs according to the system variables employed in the design of the operator. Likewise, the distribution of changes in the values of these characteristics as a function of time during the movement and the relationships between successive movements and coincident movements depend on operator design variables.

Like the design of the input, the design of the operator has been given little attention. Separate discussion of the variables is not warranted, but the listing of some of those which have received at least exploratory investigation may be profitable. They are: the bodily member (or members) employed in operating the control device; the operating position, the point of support, the degree of flexion, and the plane of action of the bodily member used; the kind of movement, whether translatory, rotary, or static, discrete or continuous, cyclic or irregular; the direction of the movement relative to the operator, to indicator movements, and to movements of other bodily members coincidently employed. Brown and Jenkins (**2**) have classified movements and outlined a program of research designed to elicit much of the information which is needed. Henschke and Mauch (**14**), Brown *et al.* (**3**), and others have made beginnings in the direction of investigating these variables. Studies by Montpellier (**8**), Barnes *et al.* (**1**), Hill and Ellson (**10**), Taylor *et al.* (**22**), and Fenn (**13**) have contributed to the knowledge concerning the characteristic pattern of movements. But the surface has only been scratched. Ultimately the manipulation of such variables may be used in two ways: (1) to provide those movement characteristics which are desired as input to the controls, and (2) to eliminate those movement characteristics which are undesirable.

Manipulative Device Variables

The device which the operator manipulates may constitute the only unit between the operator and the output of the system. This is the case, for example, when the manipulative task of the operator consists of sliding a pointer along a scale to accomplish the required settings. More often there are a number of transmission stages between the manipulative device and the output of the system. For purposes of the present discussion the device and any succeeding stages will be considered as a unit.

There are four important operator-device relationships which may vary from system to system: (1) facilitation or inhibition of desired operator design, (2) dynamic modification of operator output, (3) provision of proprioceptive cues, and (4) transformation of operator output.

Facilitation or Inhibition of Operator Design. The relationship between the placement and motion potentialities of the device on the one hand and such operator variables as mean limb flexion, plane of action, type of movement, etc., on the other is obvious. Especially to be noted is the function of device characteristics as motion or position guides in maintaining consistency of movement patterns.

The relative magnitude of input changes and required operator movements is a function of the amplification or reduction characteristics of that part of the system which intervenes between the operator and the mechanism or process which he controls. The importance of this relationship to the maintenance of operator design may be illustrated by considering the control-indicator ratio in a compensatory tracking task. A one-to-one ratio will require very slight movements on the part of the operator in the case of small deviations of the pointer from the null position. The over-all operating relationship between input and operator output may be destroyed in this portion of the input range.

Dynamic Modification of Operator Output. This variable encompasses the load characteristics of the device, friction, inertia, and damping. Relative to this variable it may be well to point out that the choice of limb, the number of limbs, and the point of support involves these same characteristics. The development of equivalent one-handed and two-handed systems may require special attention to this relationship.

Provision of Proprioceptive Cues. Feedback inputs from the manipulative device were mentioned in connection with input variables. This variable is closely related not only to load characteristics as mentioned above, but also the operating range of movement variables. The relative magnitude of input dimensions and movement dimensions may be such that the adjustment required for some inputs involves changes in position, rate, or acceleration which are too small to provide discriminable cues.

operator must position the handle at zero, to match a constant target rate he must achieve and hold the correct handle displacement from zero, etc. Acceleration tracking equipment has also been developed. Table III shows the relationships between certain input characteristics and the correlated response characteristics for three different tracking techniques.

Table III is based on constant movement dimensions, although most target courses have constant dimensions only momentarily. The implications in these relationships with respect to the role of the operator are readily discernible if the tracking task is analyzed. We will first consider pursuit tracking of a

TABLE III

TARGET MOVEMENT—HANDLE MOVEMENT RELATIONSHIPS FOR THREE DIFFERENT TRACKING TECHNIQUES

	Constant Target Position	Constant Target Rate	Constant Target Acceleration
Position tracking	Constant handle position	Constant handle rate	Constant handle acceleration
Rate tracking	Constant handle position at zero	Constant handle position	Constant handle rate
Acceleration tracking	Constant handle position at zero	Constant handle position at zero	Constant handle position

Transformation of Operator Output. The manipulative device and its succeeding stages may amplify or reduce the operator's movement to a system output movement of the same form. His movements also may be transformed into movements of a different form, voltages, pneumatic pressures, etc. It is also possible to transform his movements in such a way that they are not direct counterparts of inputs presented in the form of movement. In rate tracking the position of the control handle imparts a constant velocity to the reticle. A change of direction is accomplished by moving the handle across a zero position. Thus, reticle rate corresponds to handle position and reticle acceleration corresponds to handle rate. To match a constant position of the target the

variable speed course in one dimension with constant acceleration. (Restriction to one dimension is a common feature of experimental designs although such a restriction prevents the operator from "anticipating" acceleration changes.) Because of reaction time, errors will immediately develop even if the task is started with reticle and target aligned. These errors are errors in acceleration, rate, and position. Perfect correction of all three types of error simultaneously will bring the reticle back on target. Corrections are made by acceleration adjustments of the reticle movement. The development of further error is suppressed by the maintenance of the correct acceleration. The column on the right in Table III indicates the handle-movement dimension in which precision is necessary for each of the three techniques.

Next we will consider the same task with a constant speed course. This is another common experimental design. Errors in rate and position develop. Perfect correction of both types of error simultaneously will bring the reticle and target back into alignment. Just as in the other case, corrections are made by acceleration adjustments of the reticle movement. However, the development of error is now suppressed by the maintenance of the correct rate. The center column in Table III indicates for each tracking technique the handle movement dimensions in which precision is required.

A brief comment on error correction will serve to illustrate another aspect of the operator's role. In the constant-speed tracking task just considered, the reduction of a position error is necessarily accompanied by a temporary rate error. The continuous change in target position requires that the operator reduce both position and rate errors simultaneously in order to achieve perfect tracking. However, experimental observations indicate that the two types of error are corrected alternately. A number of investigators using constant speed courses note a tendency for operators to "tolerate" small constant errors in position rather than abandon their successful rate adjustment.

Operational Analysis and System Variables

Operational analysis is a well worked out, mathematical technique for analyzing the characteristics of physical transmission systems. It is widely used by engineers to predict the response of servomechanisms and electrical networks which are too complex to allow the computation of the performance of the whole from performance of individual components. The relationships between input and output are the basic functions, and they may be investigated without reference to intervening mechanisms.

Perhaps unfortunately, the power of these methods is restricted, for the present at least, to the analysis of linear systems only. The application of these methods, therefore, to the analysis and synthesis of closed cycle systems containing a human operator as one of the elements will be fruitful to the extent that these systems are linear or can be made linear and to the extent that the human operator is linear or can be made linear.

The concept of linearity as applied to transmission systems includes three fundamental properties. These are:

1. The normal response is a linear function of the input, in the mathematical sense.
2. The normal response at any time depends only on the past values of the input.
3. The normal response is independent of the time origin.

The application of operational analysis to human motor responses has been discussed by Ellson (**9**). Using this technique, experimental investigation of the linearity of systems containing human elements is now in progress at several different laboratories (**11, 20, 23**). The test for linearity rests on the fact that in the case of truly linear systems, if the response of the system to certain characteristic inputs is known, the response to any other arbitrary input can be determined. The test inputs consist of either a step function input (the sudden displacement of the variable preceded and followed by a steady state) or sine wave inputs over a limited range of frequencies. Successful prediction of the output of the system to a complex input from a knowledge of the system output to each of its simple components constitutes an important indication that the system is linear. The preceding discussion of system variables is directly related to the determination of the effective input in such tests and to the establishment of similarities between a system which has been tested and other systems about which information is desired.

Summary

An examination of the literature on studies related to the problems of control design yields a heterogeneous mass of unrelated facts that have been gleaned from experience and experiment with specific machines and particular operations. Little generalization is possible. This condition is largely due to inadequacies in the approaches to the problems of equipment design in general. Traditional attempts to link a better man to the machine and, through changes in existing equipment, to provide a better link between man and machine, have masked the fact that the operator and his controls are transmission elements in a control system, that the control task is performed by the control system as a whole, and that the design of its elements must logically be conceived in terms of the design of the system. The relationships between facts obtained from different systems and control situations can be understood only after system differences are clarified.

In attempting a systematic approach from this point of view we have first classified control design problems according to their source and control tasks according to their sensitivity to the transmission characteristics of the operator. Problems arising from the suitability of the transmission characteristics of control systems and control tasks in which operator characteristics play a critical role were selected for further analysis.

An evaluation of the information currently available to the designer who attempts to include the operator in the original design layout leads to the conclusion that a different sort of information is needed than that which has been provided thus far. The designer needs to know the transmission characteristics of the operator, his operating ranges, and his susceptibility to expected operating conditions. The designer especially needs to be aware of the fact that "the operator" as an element in the design of the system is an abstraction from the operator who, as an individual, is the embodiment of a multiplicity of transmission systems. The details of his role thus differ from system to system and, to a large extent, determine the classification of the system. It is necessary to analyze the ways in which his roles may differ before effective use can be made of his characteristics, before system differences may be accurately appraised, and before results of system modifications may be properly interpreted. The differences between systems with respect to the role of the operator may be considered separately as input variables, operator variables, and manipulative device variables, but their interrelationships must be considered in any research or practical application.

Research in the future should be oriented toward providing answers to the following questions: What are the transmission possibilities of the operator? What are his transmission possibilities and characteristics in each of his various roles? How best may his transmission possibilities and characteristics be exploited for a given type of control task? Operational analysis, the technique which has proved so useful to the designers of complex transmission systems and servomechanisms, may be a ready-made method for use in the search for the necessary answers.

References

1. Barnes, R. M. *Motion and time study.* London: John Wiley & Sons, Inc., 1946.
2. Brown, J. S., & Jenkins, W. O. A preliminary analysis of human motor abilities in relation to the design of equipment and a suggested program of research. In: P. M. Fitts (Ed.), *Psychological Research on Equipment Design*, A.A.F. Aviation Psychology Program Research Reports, 1947, **19**, 35–63.
3. Brown, J. S., Knauft, E. B., & Rosenbaum, G. The accuracy of positioning reactions as a function of their direction and extent. *Amer. J. Psychol.*, 1948, **41**, 167–182.
4. Channell, R. C., & Tolcott, M. A. *The use of human engineering data in equipment design problems.* Office of Naval Research, Special Devices Center, Navy Department, Contract N6-ori-151, Project 20-F-2, Report 151-1-16, 20 May 1948.
5. Chapanis, A., Garner, W. R., Morgan, C. T., & Sanford, F. H. *Lectures on men and machines.* Office of Naval Research, Special

Devices Center, Navy Department, Contract N5-ori-166, Report 166–1–19, 1947. (RESTRICTED)

6. Craik, K. J. W. Theory of the human operator in control systems, Part I. The operator as an engineering system. *Brit. J. Psychol.*, 1947, **38**, 56–61.
7. Craik, K. J. W. Theory of the human operator in control systems, Part II. Man as an element in a control system. *Brit. J. Psychol.*, 1948, **38**, 142–148.
8. De Montpellier, G. Note on acceleration in voluntary hand movements. *Arch. Psychol., Genève*, 1937, **26**, 181–197.
9. Ellson, D. G. The application of operational analysis to human motor behavior. *Psychol. Rev.*, 1949, **56**, 9–17.
10. Ellson, D. G., & Hill, H. *The interaction of responses to step function stimuli:* I. *Opposed steps of constant amplitude.* A.A.F., Air Materiel Command, Wright Field, Dayton, Ohio, 1948, Serial No. MCREXD-694-2P.
11. Ellson, D. G., Craig, D. R., & Hill, H. *The interaction of responses to step function stimuli:* II. *Equal opposed steps of varying amplitude.* A.A.F., Air Materiel Command, Wright Field, Dayton, Ohio, 1948. (In press)
12. Ellson, D. G., & Wheeler, L. *The range effect.* A.A.F., Air Materiel Command, Wright Field, Dayton, Ohio, 1948. (In press)
13. Fenn, W. O. The mechanics of muscular contraction in man. *J. appl. Physics*, 1938, **9**, 165–177.
14. Henschke, U. K., & Mauch, H. A. *A study of the design for a three-dimensional hand control for aircraft.* AAF, Air Materiel Command Engineering Division, Aero-Medical Laboratory, Wright Field, Memorandum Report TSEAA-696-110, 12 May 1947.
15. Hick, W. E. Discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, 1948, **1**, 36–51.
16. James, H. M., Nichols, N. B., & Phillips, R. S. *Theory of servomechanisms.* New York: McGraw-Hill, 1947.
17. Jenkins, W. O. *A review of psychological studies pertaining to the design of gunnery equipment.* A.A.F., Air Materiel Command, Wright Field, Dayton, Ohio, Contract # W 33–038 ac-13968, Report No. 3, March, 1947.
18. Lawrence, M., & Macmillan, J. W. *Annotated bibliography on human factors in engineering design.* Aviation Branch, Research Division, Bureau of Medicine and Surgery, Navy Department, February, 1946. (RESTRICTED)
19. Orlansky, J. *The human factor in the design of stick and rudder controls for aircraft.* Office of Naval Research, Special Devices Center, Navy Department, Project No. 20-M-1c, Report #151–1–8, 3 February, 1948.
20. Searle, L. V., & Taylor, F. V. Studies of tracking behavior. I. Rate and time characteristics of simple corrective movements. *J. exp. Psychol.*, 1948, **38**, 615–631.
21. Smith, K. U. *Psychotechnological factors in the design of flexible gunnery equipment.* Applied Psychology Panel, NDRC, AC-94, OSRD Report No. 6222, Memorandum No. 6, October, 1945. (RESTRICTED)
22. Taylor, F. V., & Birmingham, H. P. Studies of tracking behavior. II. The acceleration pattern of quick manual corrective responses. *J. exp. Psychol.*, 1948, **38**, 783–795.
23. Tustin, A. The nature of the operator's response in manual control and its implications for controller design. *J. Inst. elec. Engineers*, Part IIA, No. 2, May 1947.

Chapter 6

HUMAN FACTORS IN PANEL DESIGN

ELIOT STELLAR

The Johns Hopkins University

The proper arrangement of panel layouts is one of the basic problems in the design of military equipment for efficient human use. It has already been shown that the efficiency of performance in man-machine situations is greatly increased if individual controls and displays are designed to suit the capacities of the operator as well as the electronic and mechanical ends they serve. But even with ideally designed controls and displays, it is obvious that a great deal will depend on the arrangement of equipment panels. Both industrial and military investigators have made it clear that some arrangements of controls and displays are easy to operate and others conducive to errors and fatigue. In fact, they have started to collect data on the human perceptual and motor capacities that are directly relevant to panel layout problems and have a number of general principles to recommend for improvements in the design of equipment panels. It is the purpose of this chapter to summarize and evaluate the results of these investigations and to describe what the study of human factors in equipment design has to offer in the solution of panel layout problems.

Fact and Theory in Panel Design. At the outset, it should be emphasized that only a modest amount of headway has actually been made on panel design problems up to the present time. Although we do have satisfactory data on many questions, there are great gaps in our knowledge which have been filled in, rather liberally, with opinions and hypotheses. In fact, it is particularly worth noting that there has been a marked tendency in the literature in this area to formulate general principles of panel design prematurely and to offer them as recommendations to the design engineer. Ideally, these general principles should be the conclusions of experimental studies of human capacity in the operation of equipment. Actually, it turns out that many of them are little more than the common sense opinions of experienced engineers and psychologists. It is unfortunate that little attempt has been made to separate fact from opinion in writing about panel design. As a result, it is often difficult to distinguish what we actually know from what somebody thinks or would like to recommend on the basis of practical experience. Of course, both fact and theory are important in the solution of practical problems and in the direction of future research. But at this stage of our knowledge, it is crucial that opinion be recognized as such, and wherever possible, be subjected to direct experimental test.

Principles of Panel Layout. With this fundamental point in mind, we can now consider what has been done on the problem of panel layout. For convenient discussion, we can begin with a brief summary of the basic principles of panel layout which may be found in the literature. Actually, numerous recommendations for panel design have been made in one form or another, but they may be synthesized into six general statements which cover the major factors that have been investigated. (1) All controls and displays should be located within a practical working distance from the operator. (2) Within the practical working area, each control and display should be located in a position where it may be most efficiently used. (3) Where two or more controls or displays on a panel have the same optimal position, the more important and more frequently used items should be given the preferred positions. (4) Controls and displays should be grouped in patterns that make for the easiest operation and observation from the point of view of the operator. (5) No one part of the body should be overloaded with work that could be assigned to other body parts. (6) Any possible confusion of controls or of displays by the operator should be avoided by proper design or placement.

It should be pointed out that these principles are merely convenient devices for summarizing the best of what we know about

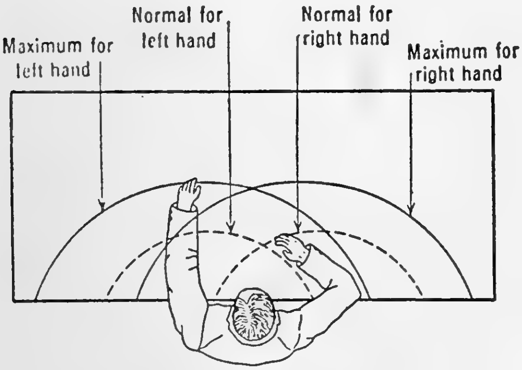

Fig. 1. Normal and maximum working areas. (From Barnes, **1**)

panel layout. How well they can be used in panel design, of course, depends on how well they are substantiated and implemented by experimental data. Therefore, it will be important to consider the experimental work on which they are based in some detail. After that, the problems involved in applying our research knowledge and general principles to practical design problems and the methods of evaluating the suitability of panels for human use can be discussed.

Studies of the Human Factor in Panel Design

The Practical Limits of the Work Space

Although it is obvious that an operator must be able to reach controls and perceive displays before he can use them, several practical problems arise in the design of equipment to conform with this principle. What are the practical limits of reach and vision of the operators who are likely to use the equipment? How important is the position the operator takes in front of the equipment in determining what his effective working area will be? How should a panel be oriented in space for the greatest possible working space and most efficient performance? Several studies bear directly on these questions and contribute the kind of specific information that is needed to design panels properly. Basically, they have all been guided by the belief that the less the operator has to move his body from resting position to use controls and displays, the less fatigue he will suffer and the more efficient will be his performance.

The Limits of Reach

Several approaches have been taken to the problem of the limits of reach of an operator. Early industrial investigators made empirical determinations of what they called the "normal" and "maximal" working areas of typical male operators sitting at horizontal surfaces (**1, 2**). They simply had subjects draw arcs on the surface with each hand, once with the elbow as the point of pivot ("normal" area) and once with the arm extended and the shoulder as the pivot ("maximal" area). This technique shows that the working areas of an operator's hands are semi-circular, and it clearly indicates the areas in which the two hands can reach equally well (Fig. 1). It is a conservative estimate of the limits of reach, however, for the limits it calls maximal are determined without any trunk or body movement on the operator's part. Clearly the reach could be extended beyond these points by merely moving the back away from the chair and changing the shoulder pivot point. Nevertheless, the technique is a simple one that could be used with appropriate modifications for a provisional determination of the limits of reach in any panel or for that matter in three dimensions (Fig. 2).

The Use of Anthropometric Data. A second, and perhaps more valuable, method of determining the limits of reach is to make use of the data on body size collected in anthropological studies. If the arm length, leg length, and standing and sitting heights of the men who are likely to use a particular piece of equipment are known, it is quite possible to calculate what distances they will be able to reach on horizontal and vertical surfaces. In one study in the Army Air Forces, such measurements were made on about 3000 cadets (Table I), and the distributions of data were presented in such a way that it was possible to tell what distances would be beyond the reach of 5%, 50%, and 95% of the population if the body were held in a fixed position (**31**). In a similar kind of investigation, in the Navy, even more direct data on the limits of reach in the cockpit were obtained (**21**). In this case, only 139 subjects were used, but the distributions of measurements obtained agreed fairly well with those of the larger cadet sample. All measurements were made with the subjects seated properly in a standard pilot's chair and extending their right arms to reach horizontal and vertical measuring rods which gave coordinates from a reference point on the seat (Fig. 3). Analysis of the data showed the maximal horizontal and vertical distances that could be reached by 97% of the population when the body was fixed against the chair back and the chair tilted 13° backwards (Figs. 3 and 4). In both of these studies, again, the estimates of the working area are conservative, for all measurements were made with the subjects' heads and bodies in fixed positions. Even though it is considered desirable to keep gross trunk movements and stretching at a minimum, it would be important to know how far an operator could reach, if he had to, and still operate his equipment.

Both of these techniques were applied in a recent study of the limits of reach on a vertical surface (**24**). Eight subjects were seated at viewing distances of 10, 15, and 20 inches from a wall chart on which they traced the movements of their outstretched hands while their heads and bodies were fixed against a chair back. By comparing the body sizes of these subjects to the population of aviation cadets, the limits of reach of 5%, 50% and 95% of that population could be calculated. Fig. 5 shows a sample of the detailed information that this type of study yields. The two hands traced overlapping circles that varied in size as a function of the physical size of the operator and the distance he sat from the vertical surface. For example, it was shown that to include 95% of the population, at a viewing distance of 20 inches, the

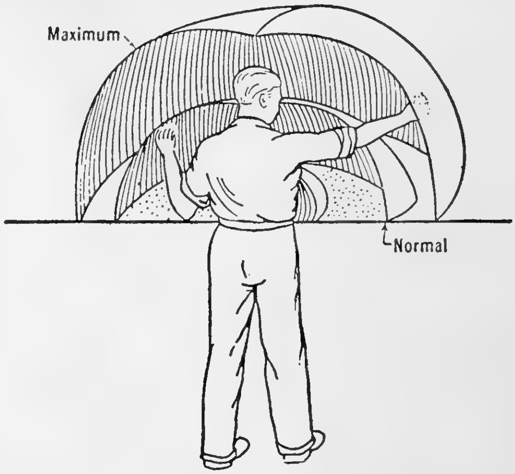

Fig. 2. Normal and maximum working space in three dimensions. (From Barnes, **2**)

TABLE I

SOME BODY SIZE MEASUREMENTS OF ABOUT 2960 AVIATION CADETS, SHOWING THE 5TH, 50TH, AND 95TH PERCENTILES (IN INCHES) OF THE DISTRIBUTIONS

(From Randall *et al.*, **31**)

Measurement	Percentiles		
	5th	50th	95th
Standing Height	65.4	69.2	73.1
Sitting Height	34.5	36.5	38.5
Anterior Arm Reach	32.7	35.2	37.8
Buttock to Knee	22.0	23.6	25.6
Patella to Knee	20.4	22.0	23.6

limits of reach would have to be confined to the two overlapping circles making up an elliptical area, about 43.9 by 34.8 inches while an area of 60.6 by 51.5 inches could be reached by the largest 5% of the population. At a viewing distance of 10 inches, the values for 95% and 5% of the population were 58.1 by 49.0 inches and 71.1 by 62.0 inches, respectively. This type of information should be invaluable for planning the practical limits of the working area on a vertical surface for operators of different body size and at different distances from the panel. The same kind of study should be done with other types of panels and work areas now in current use.

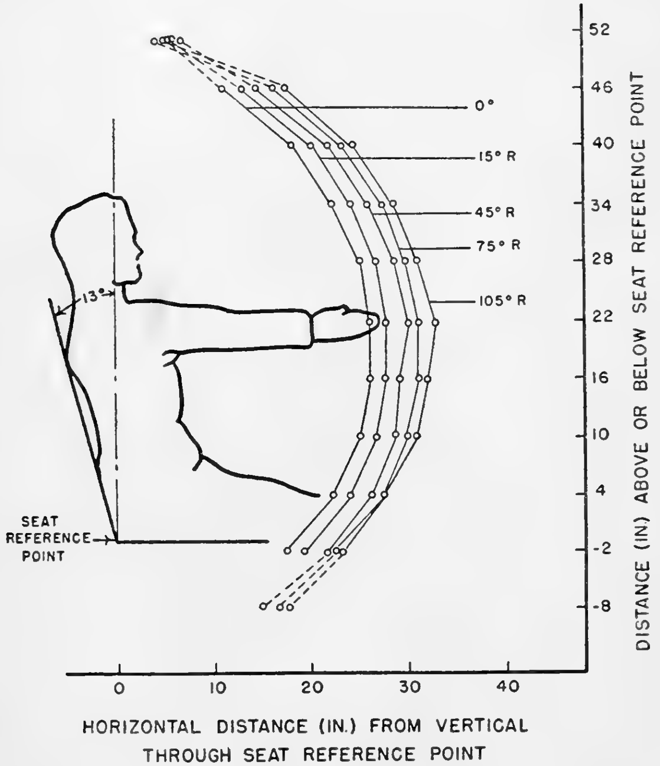

Fig. 3. Maximum distances which can be reached by 97% of a selected Navy population at angles from 0 degrees to 15 degrees to the right of the midline. (After King *et al.*, **21**)

In actual practice, it has been shown that, in many types of equipment, controls are not within the limits of reach unless the operator stretches and moves his body. In one investigation of multi-engine aircraft, the R5D (Navy equivalent of the DC 4), it turned out that most of the controls were beyond the reach of the pilots if they were required to keep their bodies resting against the back of their seats (**7**). Measurements from the standard reference line on the seat used by King to the midpoints of seven control areas of the cockpit showed that all but one area would be out of the limits of reach of 50% of the population of the subjects used in that study (Table II). Furthermore, photographic records of flight operations with this aircraft showed that pilots actually did a great deal of trunk twisting and stretching to reach controls. In another study of errors made by pilots in flight, it was found that 3% of the errors which they could remember were due to the inability to reach a control at all (**13**). In most of these cases, the control was not beyond the limits of reach if the pilot stretched from the resting position, but rather were out of reach because the pilot had to stretch in the opposite direction to operate another control at the same time.

Taken together, these studies offer the kind of specific data that are necessary for efficient panel design. In the first place, it should be obvious that we can now determine the limits of reach of any population in any kind of equipment panel without resorting to guesswork. With the techniques available, it is possible to specify precisely what areas of a panel a particular percentage of the population can reach without stretching or straining. There should be no excuse for unforeseen difficulties on this score in

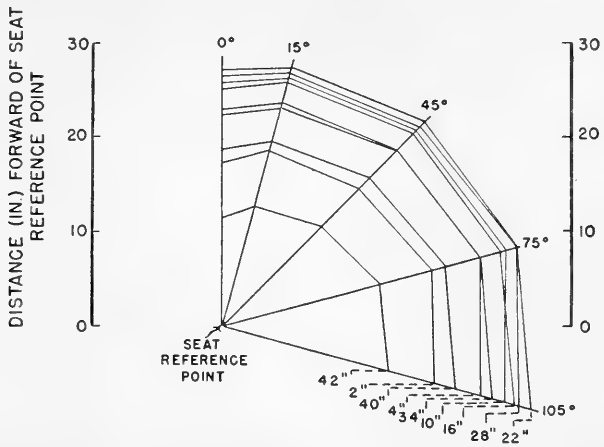

Fig. 4. Maximum distances which can be reached with the right hand by 97% of a selected Navy population at vertical intervals from 46″ above to 2″ below seat level. (After King *et al.*, **21**)

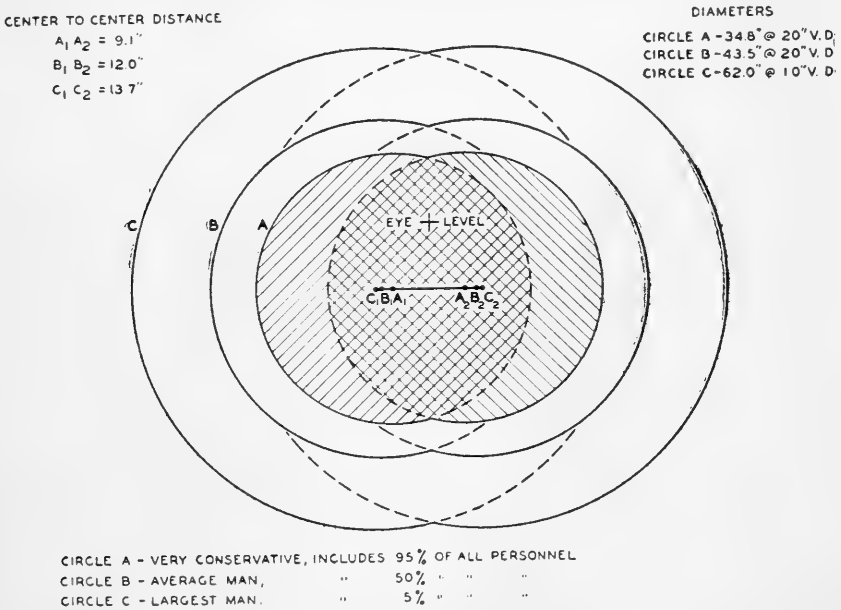

Fig. 5. The limits of maximum working areas on a vertical surface. (After Lipschultz and Sandburg, **24**)

new types of equipment. Wherever possible, old equipment should be revised to meet the requirements set by the limits of reach of the population using it. At the present time, however, there is still no answer to the question of how detrimental to performance stretching to reach controls may be. In some equipment, overcrowding of panels makes it impractical to place all controls within easy reach of the operator. It would be valuable to know, therefore, what amount and what kind of stretching

TABLE II

A Comparison of the Distance (in Inches) Required to Reach Areas of the R5D Cockpit with the Maximum Reaches Possible for 50% of King's Navy Population

(From Channell, **7**)

Cockpit Area*	Distance to reach area midpoints	Maximum Reach for King's Population	
		50%	95%
2	36	32.0	27.5
3	27	25.0	20.0
4	41	30.0	25.7
5	38	31.3	27.0
6	26	17.0	11.0
7	23	30.5	27.0

* See Fig. 10 for diagram of cockpit areas.

can be tolerated without introducing enough fatigue or strain to impair normally required performance. Of course, a great deal will depend upon the nature of operation and frequency of use of the controls which the operator might have to strain to reach.

The Practical Working Area for Displays. The limits within which visual displays may be placed in the working area are a less clearly defined problem than the one faced in the case of controls. A great deal depends upon the general illumination which can be tolerated on the panel, the size, shape, and detailed design of the displays, and other demands which are made on the eyes of the operator. As in the case of controls, it has been a matter of general belief that fatigue is least and performance best when there are a minimum of eye and head movements required in the viewing of displays. Under the most extreme conditions of head and eye movement, with the body fixed in position, the average operator can fixate his eyes from 90 degrees above the horizontal to about 90 degrees below; and at least 90 degrees to either side (**31**). With unlimited head and body movement allowed, of course, vision in almost any angle around the operator can be achieved. In the other extreme, where the head is held in position and the eyes are fixated on a point in front of the observer, only a relatively small area around the point of fixation is in clear focus. For one eye, vision at five degrees from the fovea is only about one half as good as central vision (**9**). The value is higher for the two eyes together, but in any practical case depends upon the size of the objects to be seen, the distance from the eyes, general illumination, etc. In actual practice, of course, maintenance of steady fixation is rare; both the eyeballs and head rotate in the process of fixating objects out in the periphery. There are two basic problems in determining the practical limits of vision, then: (1) How much do the fatigue and time involved in making head and eye movements impair performance? (2) At what distances and angles around an observer may displays be placed without impairment of good detailed vision? As far as this second point is concerned, it is obvious that a display that is placed far out laterally on a vertical surface cannot be viewed squarely and therefore cannot be read accurately. But we have no data at present to tell us how far laterally a display may be placed on vertical, or any other type of surface and still be seen clearly enough to be used. Before we can be specific about the practical limits of vision on display panels, answers to both of these questions must be worked out experimentally. Until that time, common sense, hit-or-miss methods are the only ones at our disposal.

The Position of the Operator

The spatial limits within which the operator can work his equipment depend a great deal upon the position which he assumes in front of the panel. In determining the limits of reach and vision in the above experiments, the position of the subject's body was always fixed, his shoulders and hips in the case of hand and foot reaching and his eyes in the case of vision. Any deviation from these fixed positions, of course, throws all determinations off by some constant amount and might, for example, put some controls out of his reach. With proper panel design, the operator should be able to get into a position to reach all controls by placing himself at a suitable distance from some standard control and display reference points. Since a man has to be comfortably seated for long periods of time in many types of equipment, however, he should be able to move his seat with him as he has to change his position. One investigation, relevant to this problem, is a study of the proper seating of 52 subjects who corresponded in body size to the 3000 aviation cadets measured in the anthropological studies (**29**). While the study was rather specific to cockpit problems, it is of some value to discuss it here, for it represents a good model of the kind of work that has to be done in connection with all seating problems. Each subject was allowed to make his own seat adjustment in a cockpit mock-up, and then measurements of the distances to the various parts of the cockpit mock-up were made. Several recommendations were made on the basis of the results that were found. First, since the range of distances from the reference point of the seat to the rudder pedal was 35.5 to 41.5 inches (mean 39.1 inches), it was concluded that cockpit seats should have six-inch fore and aft adjustment that can be made in one-inch increments. Second, the range of seat to eye heights was 28.1 to 35.1 inches (mean, 31.8 inches) with the chair at 90 degrees; with the chair at 103.5 degrees, the range was reduced to 26.9 to 34.0 inches; natural slumping lost another 0.3 to 2.6 inches in seat to eye height. In any case, it seems desirable to have the height of the seat sufficiently adjustable to be certain that most operators will be able to get into the optimal position and still be comfortable.

Seating Comfort. In addition to the problem of positioning operators so they can reach controls, some details have been worked out on what constitutes the most comfortable seating arrangement in the cockpit type of seat. In one study, the angle of flexion of the arms and legs positioned on the rudder and stick were varied by adjusting the distance of the seat from these controls (**17**). No difference was found in the performance of a pursuit task when angles of 105, 120, and 135 degrees were used for either the arms or legs. But a survey of the opinion of the subjects showed that a 120 degree angle of the leg and 105 and 120 degree angles of the arms were judged the most comfortable. Within certain practical limits, then, it appears that the operator can select the most comfortable working position without a decrement in his performance. Under some circumstances, however, individual adjustment is apparently not desirable. When 20 pilots were allowed to choose their own seating arrangements in a standard cockpit, it was found that they could sit comfortably for an average time of four hours and nineteen minutes (**30**). Then, when 21 other pilots were required to sit in the average seating arrangement used in the first experiment, it turned out that they could sit comfortably for an average of seven hours. Apparently, there are certain limitations in the choice of individual pilots, so that it is better to determine experimentally the most comfortable seat for long sessions of operation than to have operators choose their own.

Other Seating Problems. More complete data are needed on the role of seating in the efficiency of operating panels before we can know which arrangements are the most comfortable and permit the easiest operation of the equipment. The Air Force studies are good models, but studies of other types of equipment and other types of operator position problems are necessary. For example, it may be desirable, as has been found the case in industry, to have the operator shift from seated to standing positions when re-

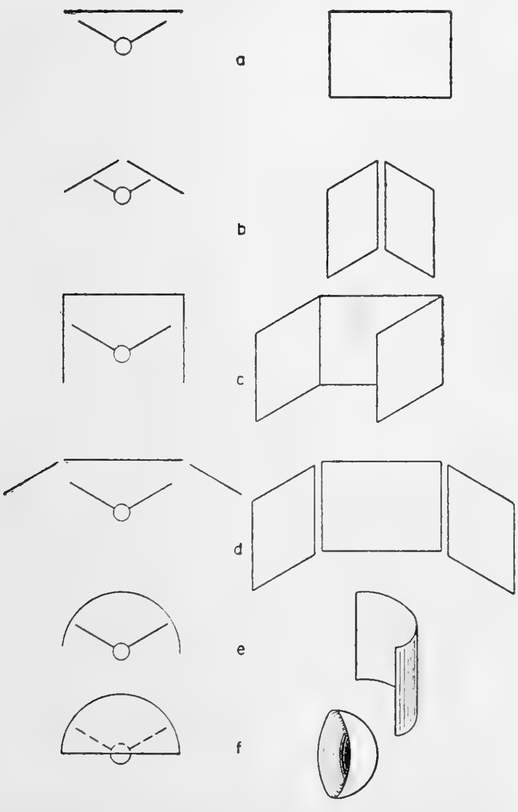

Fig. 6. Suggested types of vertical working surfaces that might increase the area an operator can reach on a panel. (From Lipschultz and Sandburg, **24**)

quired to operate equipment for long sessions (**1**). In this case, it has been found possible to arrange seat height and work space dimensions so that the operator may always be within practical working distance of his equipment and perform his duties equally well in either position. In other instances, special problems may arise with certain equipment that will require some departure from the conventional positions of the operator. One extreme example of this type is the investigation of flying in the prone position. So far it has been shown that subjects may use the best designed prone bed for as long as two hours without notable discomfort (**10**). Of course, the limits of reach and vision are seriously altered by the assumption of the prone position. Controls and displays above the head and toward the rear of the operator are extremely difficult to use; forward locations and those on the floor are the best (**5**).

The Orientation of the Panel

What position the operator will assume in the use of equipment and what limits of reach and vision he will have will, of course, depend a great deal upon how the panel or panels he is operating are oriented in space. A panel may be a single surface or a group of surfaces and may be oriented horizontally, vertically or obliquely. Which type of orientation will be used in any specific situation will depend to a large extent upon the amount of space available for the equipment. Certain conditions will demand a vertical panel, others will permit horizontal or oblique panels. Very little information is available at the present time to indicate which type of orientation best suits human capacity in different types of tasks. One study of the VJ radar mounted at 0, 15, 30, 45, 60, 75, and 90 degrees showed that the speed and accuracy of target indication was not affected by angle of inclination (**23**). Interestingly enough, however, a survey of the opinion of the experienced operators used as subjects showed that 15, 45, 30, 60, 0, 75, and 90 degrees were preferred, in that order, even though zero degrees is the normal working angle for that radar equipment. Another study of the limits of reach on a vertical panel suggests that a greater percentage of the possible reach of an operator can be utilized if panels are flanged outward toward the operator or curved or bowl-shaped (**24**). Some of the advantages achieved by these innovations are illustrated geometrically in Fig. 6. At the present time, however, there are no experimental data to indicate whether such arrangements can be practically employed in equipment design. It may be that, as equipment gets more complex and more space for controls and displays are needed, several panels placed around a single operator, or two operators working as a team, will have to be used. This is the practice in many current types of aircraft. But whether it is the best arrangement for all types of situations, only direct experimental attack on the problem will tell.

Optimal Location of Controls and Displays

Within the practical limits of the working area, controls and displays should be placed in optimal positions for use by the operator. The basic questions to be asked here are: What positions on the panel are best for the use of different types of controls and displays? Where should a control best be placed which demands strength, speed, or precision in operation? What should be the position of displays that have to be read accurately and those that require only a cursory check? In general, it is pointed out by Bartlett that the best position for most hand operated controls is directly in front of the operator between shoulder and waist height (**3, 4**). Similarly it may be said that displays placed at eye height, in front of the operator will be seen most accurately and with least effort. In fact, one study of the placement of alternative displays in this position on the instrument panel showed that pilots tended to fixate the displays in the central area the largest number of times regardless of whether they indicated horizon, direction, or turn and bank in flight (**20**). Of course, one might well expect that other positions on the panel will serve as well or better than these central ones for the use of certain controls and displays.

Localization of Controls

As far as simply reaching accurately for controls is concerned, Fitts has shown that performance is best when they are located at shoulder height in front of the operator (**12**). He studied the ability of 50 pilots to localize targets placed around them in a cockpit mock-up. The subjects had to fixate a red bulb through slits in opaque goggles, thereby fixing the head in simulation of certain flight conditions. Then, in response to instructions they tried to locate the center of each target by marking it with a small stylus held in either the left or the right hand. When required to reach for targets above shoulder level, it was found that they tended to aim too low, and with targets below shoulder level, too high. Errors in localization increased with targets placed on either side of the operators and became largest with the targets located farthest to the rear.

Placement of Controls

Very few experiments have been done to carry the problem of the placement of controls beyond studying the accuracy of reaching for them. There is one investigation of the efficiency of turning handwheels as a function of their location on a $47\frac{1}{2}$ by $47\frac{1}{2}$ inch vertical surface (**35**). It showed that there were limited areas within which wheels of $1\frac{1}{2}$ inch diameter could be operated most efficiently by each hand. For 11 seated subjects, the best region for each hand was an irregularly shaped area on the panel, about 16 by 20 inches, with its center about 12 inches below the horizontal line of sight and 20 inches to the left and right of the midline of the body for the left and right hands, respectively. The maximal areas within which the handwheels could be operated with 80% efficiency by each hand were more than twice the size of the optimal areas and were well within the limits of reach of the particular subjects used. Some information is available on the placement of hand controls for the maximum amount of pull and push (**19**). In general, pull is maximal when the hand grip is at about elbow height, and the subject has his seat far enough from the control so that his arm is comfortably extended at the start of movement. Having both back rests and foot rests contributes to the amount of pull possible. The optimal conditions for pushing a hand control are about the same as for pulling, but in addition it was found that maximal push was greater than maximal pull and that pushing was more fatiguing subjectively. Similarly in a study of tank drivers, it was found that the greatest force could be exerted on a foot pedal when it was so located that the thigh made a 20 degree angle with the horizontal and the knee angle was 165 degrees at the end of the pedal stroke (**18**).

More detailed information of this sort is needed about the optimal positions for the operation of other types of controls and for the placement of displays in the best positions for use. In many studies, certain design specifications for controls and displays are recommended as the best for human use, but all too often, the experiments were done with the controls and displays directly in front of the subjects. One might ask whether the design specifications hold well for other locations in respect to the operator or whether a different type of design would be more suitable if the item were to be placed in the periphery of the panel, for example. At least it would be important to know in what locations, other than the central ones, each control and display may be used and how much impairment of performance results from moving them different distances from their optimal positions. That is to say, we should have data not only on the optimal position of each item on a panel, but also the second and third best positions in which they may be placed.

The Importance and Frequency of Use of Controls and Displays

Methods of Investigation

The decision to place a control or display in a position other than its optimal one depends primarily upon its importance and frequency of use. If a particular control or display is crucial to the operation of a piece of equipment, then it must be given the highest priority and placed in a position where the operator will never fail to use it readily and accurately. Usually only a limited number of items on any panel will fall into this category. For the bulk of controls and displays, the frequency of use must serve as the basis for deciding whether or not they get placed in the preferred positions. Various methods have been used to determine the importance and frequency of use of controls and displays. As far as importance is concerned, the only way to determine that is to get the opinion of experienced operators or the engineers who built the equipment, by asking them to rank order or rate each control or display according to how crucial it is in the operation of the equipment. The same technique may be used to ascertain the frequency of use of controls and displays, but there are two other methods which perhaps yield more detailed information on this question. One that may be used in studying controls is the micromotion analysis technique employed in industry. In this method, motion pictures are taken of samples of the work of experienced operators. By examining the developed film, it is possible to count accurately how often each control is used in different types of operations. In investigating the frequency of use of visual displays, records of eye movements are of considerable value, for it is possible to determine accurately the number of fixations the operator makes on each display as he operates his equipment.

Essentially these methods have been used in the analysis of the importance and frequency of use of controls and displays on one type of radar console (**34**). Here the subjects were asked to rate each control and display as very important, moderately important, or of little importance, and frequently, occasionally, or infrequently used. Then, instead of employing micromotion techniques, the investigator observed the subjects directly during the operation of the equipment and tallied the frequency of use of each control and display. The results on frequency of use obtained with these two techniques agreed reasonably well, and together with the data on importance, provided a basis for some fairly clear cut distinctions among the different controls and displays. In another study, the method of micromotion analysis was applied successfully to the more complicated operations involved in flying multi-engine military aircraft (**7**). The problem in this case was infinitely more complex, for different controls and displays ranked higher in frequency of use in different types of operations such as takeoff, cruising, landing, etc. Nevertheless, a clear basis was provided for distinguishing among the various controls of the cockpit.

In view of the need for greater simplicity in the panels of most current military equipment, the less important and rarely used controls and displays may be located off the main surface of the panel or at least on its periphery. Several novel suggestions have been made in connection with these minor controls and displays and are worth mentioning here. For one thing, infrequently used controls, especially switches and small knobs, may be recessed into the face of the panel so that they will be out of the way, even though perhaps somewhat harder to operate. For displays used only rarely or in special operations, it has been suggested that they be of the plug-in variety so that they can be stowed away when not in use, but plugged into optimal positions when needed (**15**). Both of these recommendations are good, but should be subjected to experimental test.

Patterning of Controls and Displays

Under certain conditions of panel design, it may be possible to shorten the layout procedure by first determining the key controls and displays and placing these in positions for optimal use. Then the remainder of the items needed in the panel can be placed according to their relationship to the key items, by following the logic of arrangement of controls and displays in patterns most suited to human use. The use of the first step in this type of procedure is illustrated by an actual example of the redesign of one type of Navy aircraft instrument panel (**15**). In this case, it was considered that the basic instruments used in the "attitude" system of instrument flying, the artificial horizon and the directional gyro, should be given the central positions on the

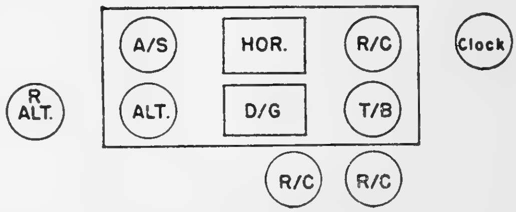

Fig. 7. Naval aircraft instrument panel designed according to the principle of functional grouping of displays. (From Foley, **15**)

panel. Once these important and frequently used displays were placed, displays of secondary importance and use associated with them were placed conveniently around them (Fig. 7). The rate of climb indicator was placed to the right of the artificial horizon because these two instruments are often used together. The air speed indicator, sensitive altimeter, and turn and bank indicator were placed close to these basic instruments in the central area because these were considered rather important adjuncts to the attitude flying instruments. Then, on the periphery of this central region, instruments were arranged near the basic displays with which they were functionally associated. The radio altimeter was placed to the left of the sensitive altimeter, and the radio compass indicator and remote compass indicator were placed below the directional gyro since both of these instruments are direction indicators.

In following this system, then, the primary decision is made in the placement of the main items in a group; then the allied items are grouped around them in a pattern that makes for easiest use by the operator. Certain practical difficulties may be encountered in following this logic, however. For example, in some models of planes, it was found that there was not enough room to place the artificial horizon above the directional gyro, since both of these are large case instruments. As an alternative arrangement for such planes, it was decided to place the artificial horizon to the right of the directional gyro. Then the rate of climb indicator and other associated instruments were relocated in the best functional relationship to the new positions of the basic instruments. In actual test, it turned out that there was very little difference in performance with these two alternative arrangements when pilots were required to fly prescribed courses on instruments (**26**). Interestingly enough, of the fourteen pilots studied in this situation, nine preferred the second panel to the first. Both of these items of information are important to know in deciding between alternative panel arrangements.

Functional Grouping

It has been pointed out frequently in the literature that both controls and displays are best used when arranged in functional groups (**11, 15, 25**). While it can be accepted as a matter of common sense that items on a panel that are used together ought to be placed together, there has been some disagreement over what makes up a functional group. It should be made clear that grouping of controls and displays according to machine function does not always make sense from the operator's point of view. It has been pointed out, for example, that pilots have to use flight instruments and engine instruments together all of the time, so that separation of engine and flight instruments would actually be detrimental to efficiency (**20**). What is needed is an analysis of the behavior of an experienced operator using his equipment. Then, when it is learned which controls and displays are used together, they can be arranged in functional groups and patterns that will be the least wasteful of motion and eye movements and the easiest for the operator to carry out.

Not too much research has been done to indicate what groupings of displays are the best for human use, but some recommendations have been made. One analysis of aircraft instruments suggests that they may be classified according to whether they are to be read quantitatively, qualitatively, or merely given a check-reading. For instruments to be check-read, it is suggested that similar purpose items may be arranged in a compact group in such a way that all the indicators face in the same direction when the instruments are set properly (**8, 25**). In general, displays that are read in sequence should be arranged so that eye movements are kept at a minimum. Similarly, related groups of displays should be located in respect to each other so that they can be observed one after the other with minimum eye and head movements. By means of eye movement studies it is possible to show the sequence which an experienced operator follows in reading instruments in normal operation. Redesign of the panel layout can then achieve the goal of minimum eye and head movements. This sort of procedure has proven helpful in industry, and can be applied successfully to tasks as complex as piloting military aircraft. More advantage should be taken of this tool in the study of optimal panel layouts.

Analysis of Links

Another procedure that might be used, not only for the arrangement of displays, but also for controls, is a technique developed for the arrangement of men and machines in a complex communications center (**9**). Basically, the method is to determine the functional links that exist among the items in the system, in this case, the controls and displays on the panel. Links may be defined in terms of both the frequency with which the use of one control or display follows another in time and the importance that may be attached to using two items in succession. This information may be obtained by surveying the opinion of experienced operators, micromotion analysis, and eye movement recording. Once it is obtained, some sense can be made of the problem of functional grouping of controls and displays by simple graphical solution. Each control and display may be depicted on a chart, and for each link it has with another control or display in actual operation, a line may be drawn from one item to the other. Then the controls and displays can be arranged on the chart so that they are closest to the other items with which they have the most frequent and important links. Finally these relative spatial relations can be used as a guide for the placement of items within the practical working limits of the panel. Up to the present, this technique has not been used in panel layout problems. However, its successful application in the arrangement of communication centers recommends it for further experimental evaluation in connection with panel layout problems (Fig. 8).

Patterns of Movement

The argument for functional grouping of controls has been made even more strongly than in the case of displays, for waste motion of the trunk and limbs is considered even more conducive to fatigue and impaired performance than waste eye movement. Within any functional group of controls, however, there is still the problem of proper placement of individual controls in the most efficient pattern for operation. On the basis of industrial studies certain types of movement are recommended as the more easily performed by humans: ballistic movements, curved movements, continuous movements, and use of the two hands simultaneously in

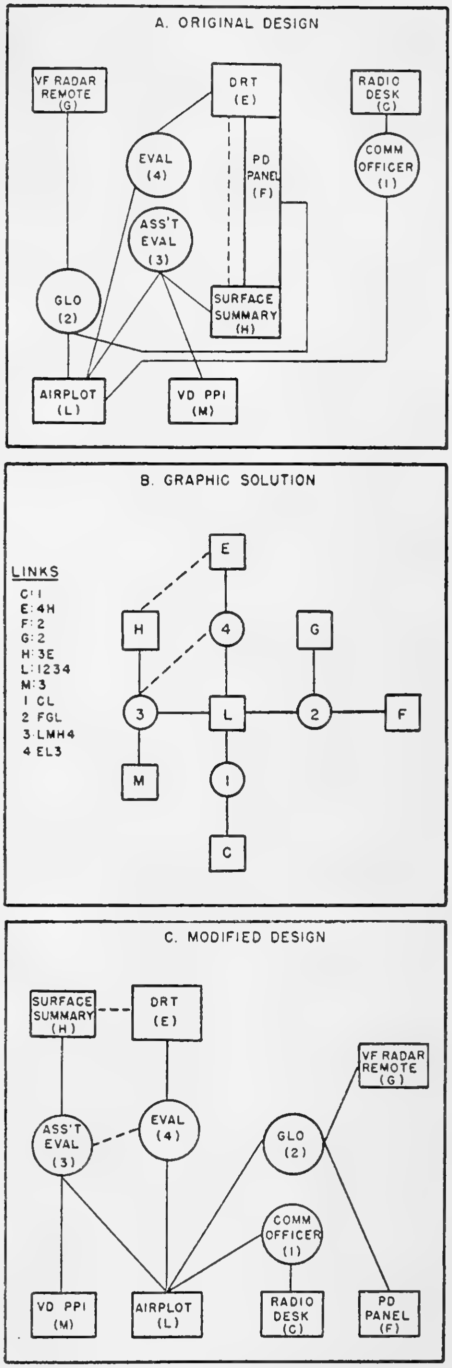

Fig. 8. Analysis of functional links in a complex man-machine situation, showing the original design of a communications center (A), the graphical solution of link values (B), and the improved design (C). (After Chapanis *et al.*, **9**)

opposite directions (**1**). In several investigations, it has been found that redesigning work-layouts to permit these types of movements has had significant results in increasing the efficiency of operation and the output of workers. For example (**27**) in a simple soldering operation, output was nearly doubled when as many of these principles of movement as possible were incorporated into the rearrangement of the items used in the operation (Fig. 9). Similar advantages could probably be gained by arranging the panels of military equipment to allow more efficient movements in the operation of controls. Only direct experimental test will tell.

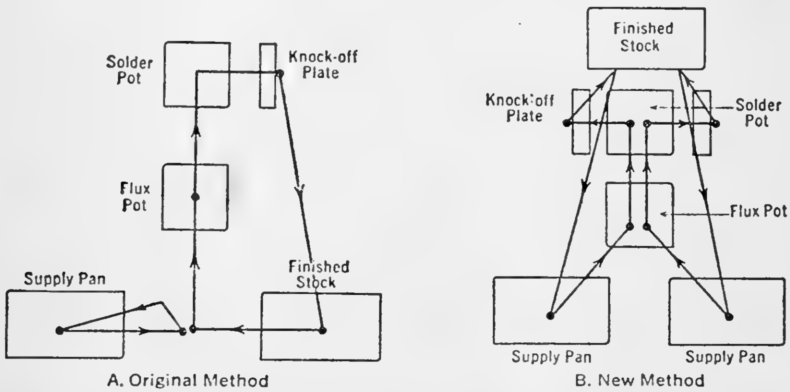

Fig. 9. Redesign of a work layout to permit the most efficient patterns of movements. (From Mogenson, 27)

Distribution of Work over the Body Parts

Part of the same problem of arrangement of controls and displays is the question of overloading parts of the body with work. In general, it is said that there should be equitable distribution of work for all parts of the body. As far as controls are concerned, it has been pointed out that the hands are usually overloaded, and that wherever possible, the feet should take over the operation of controls. There are some obvious limitations on what the feet can do. One study has shown that the feet and legs are distinctly inferior to the hands and arms in pursuit tasks (**17**). Operating either a stick or a wheel control, subjects were able to stay on a target much more of the time than they could by using rudder controls. More studies of this sort are needed, comparing the hands and feet in other military operations, for industrial investigations have shown that in many types of work, rather delicate operations can be performed with foot pedals and knee levers (**1, 2**). When the hands and arms alone are considered, it turns out that in some current equipment, the right hand is greatly overloaded while the left remains idle a great deal of the time (**2, 7**). Wherever possible, this type of situation should be rectified by relocation of controls so that they are more symmetrically distributed about the midline of the operator's body.

Classification of Movements

In keeping with a principle of least effort, it has been recommended in many industrial studies that the operation of controls be assigned to the lowest classification of body movements (**1, 2**). Finger motions are considered the easiest to make and hence given the lowest classifications; wrist, forearm, upper arm, and shoulder movements are considered more difficult in that order. This classification principle implies that up to a certain point at least, the more controls that can be operated by the fingers, the less the fatigue, and the greater the efficiency of performance will be. To a large degree, of course, whether just finger and wrist motions or motions involving the use of the shoulders will be necessary depends a great deal upon whether controls are placed within easy reach of the operator or out on the periphery of the practical working area.

Relieving the Eyes of Work

The biggest problem of this sort on the display side is the overloading of the eyes. The eyes can only do so much work in finite time, but there has been a tendency, in the design of equipment, to display most information visually. A few studies have investigated the possibilities of greater use of other sensory modalities than vision to get information to an operator. Some aviation research has been devoted to this question, for the display problem there is as complex as anywhere. Both British and American laboratories have experimented with the use of auditory signals in the attitude system of instrument flying (**6, 16**). In the American system, called Flybar (flying by auditory reference), it is recommended that turn, bank, and airspeed indications be given by auditory signals of different frequencies, intensities, and interruption rates to the two ears. In a pursuitmeter task, it has already been shown that direction can be given by such auditory signals. The auditory indication is as good as the visual when only one item of information is given; when two items are given or all three, turn, bank, and airspeed, use of the visual system is superior. More research on this problem may work out more practical methods for giving auditory information of this sort.

Confusion Errors

Since accurate and rapid identification of controls and displays on a panel is an important prerequisite for efficient performance by the human operator, it is necessary, in the design of equipment, to guard against any possible confusion of one control or display with another. In the case of many controls and displays, their general appearance, location, or mode of action are sufficiently distinctive so that confusion rarely occurs. But as panels become more crowded and the same type of knob or dial is used for more than one control or display, greater demands are made on the operator's discriminative capacities, and it becomes necessary to take special precautions to avoid the accidental use of one control or display in place of another. In any practical panel design situation, three basic problems must be investigated if confusion errors are to be reduced to a minimum: (1) finding which controls or displays actually are confused in the operation of equipment, (2) determining what conditions will make human discrimination the easiest and most accurate, and (3) planning the arrangement of panels and the design of controls and displays so that as many as possible of these conditions are satisfied. Some headway toward the solution of these problems in the cockpit has been made in recent studies in the Army Air Forces.

In one investigation of aircraft controls where 460 pilots were asked to describe one error which they remembered making in flight, it was found that about 50% of the mistakes reported could be classified as errors of confusion (**13**). Of these, 39% were made on the throttle quadrant where the throttle, propeller pitch, and gas mixture controls are located. It turns out that these three controls have different positions on the quadrant in different types of aircraft: in the B-25, the order from left to right is throttle, propeller, and mixture; in the C-47, it is propeller, throttle, and mixture; and in the C-82, mixture, throttle, propeller. The confusion in this case arises from the fact that the pilot has learned to rely on the kinesthetic cues arising from his own movements to guide him in selecting the proper control, yet the patterns of movement learned in one airplane are inappropriate in other ones. It would be like trying to drive an automobile in which the clutch and break pedals were interchanged in position—there would be many confusion errors. In the same study, it was found that 31% of the confusion errors were cases of confusing the wheel and flap controls. In many aircraft, both these controls are switches, located near each other and used together in landing and take-off operations. They look and feel so much alike that a small error in reaching can result in the activation of the incorrect control. What the pilot apparently needs is to have the two switches sufficiently far apart so that reaching errors will be rare, or to have the two controls markedly different in appearance so that he will know immediately when he is about to make a mistake.

Standardization and Coding of Controls

Two major recommendations for the design of panels are obvious from this study. In the first place, the position of all controls in different models of the same type of equipment should be standardized. Secondly, controls should be made as distinctive as possible. Several specific suggestions have been offered for making controls distinctive: (1) coding control handles by shape, color, or size so that the operator learns to associate one type of handle with each type of control, (2) spatially separating controls, that are apt to be confused, to avoid reaching errors and to afford the operator differential kinesthetic cues on which to base his discriminations, (3) providing controls with distinctive modes of operation such as up and down, left and right, in and out, etc., (4) as a last resort, using a system of warning lights or sounds to aid in the rapid rectification of errors or using mechanical guards to prevent accidental activation of controls. The general principle underlying these techniques for making controls distinctive is that the more sensory cues an operator has to go on in making discriminations, the more accurate will be his performance. Therefore, the design engineer should be advised that the more distinctive visual, tactual, or kinesthetic cues that can be built into controls, the fewer will be the errors of confusion. In fact, it has been suggested further that the kinesthetic cues that are already inherent in reaching for controls in different positions could be used more profitably if operators were trained in "blind reaching" for controls; that is, were trained to use kinesthetic cues deliberately.

Only a start has been made in the investigation of the feasibility of these recommendations. A study by Wietz has shown that shape and color coding are very effective in reducing confusion errors (**36**). He used a mock-up situation in which pilots learned to associate each one of four identical levers with four different visual signals. When the positions of the correct levers were changed, performance in this situation was greatly impaired and almost complete retraining was required. If each lever had a distinctive shape and color in the original learning, however, confusion errors were at a minimum when the positions of the correct levers were changed as long as the distinctive shape and color of each lever were still associated with the same visual signals as in original learning. Shape coding alone served fairly well to reduce confusion errors, but the effect was not as great as when both shape and color were used.

Although these results show clearly that shape and color coding or shape coding alone help to reduce confusion errors when the position of controls are interchanged, it should be pointed out that further work on this problem is necessary. In the first place, it has to be shown that coding will work as well in the operational situation as in the experimental mock-up. Secondly, we have to be certain that the shapes used for coding do not impair the efficiency of handling individual controls. Other experiments have shown that the speed and accuracy with which controls may be used and the amount of strength one may apply to them depends upon the handle or knob shape (**22, 33**).

Third, further work must be done on the use of size coding and the use of shape, color, and size coding on controls other than levers. In general it may be said that coding controls makes good psychological sense as far as the reduction of confusion errors are concerned, but it must be tried out in the practical situation where the overall efficiency of the operator can be measured with and without the benefit of coding.

Standardization and Coding of Displays

Practically no work has been done on the problem of confusing one display for another. For one thing, the problem appears to be less serious than in the case of controls, at least as far as aircraft are concerned. In a study on errors made in the use of displays, Fitts found that only 13% of the errors recalled by pilots could be classified as substitution or confusion errors (**14**). He suggests, as in the case of controls, that these errors may be reduced by standardizing the positions of displays in all planes, and by designing displays that are likely to be confused as distinctively as possible. Similar recommendations may be found elsewhere in the literature (**25**). In general, it is suggested that displays that might be confused should be separated spatially, be given distinctively different shapes, be coded by color, and possibly by size. Here again much specific research is necessary before it is possible to know how to design displays that will not be confused with each other by the operator. And again, it is necessary to be sure that making displays distinctive does not interfere with the efficient use of the individual displays themselves.

The Application of Research Knowledge to Panel Layout Problems

When it comes to the application of our knowledge of human capacities to the practical panel design situation, it is obvious that the data and recommendations considered above represent only a fraction of what we need to know. Much more experimental work is needed before we have anything like complete information relevant to problems of panel layout. The six principles described above serve well to delineate some of the major problems in panel design. As new knowledge of human capacity is applied to panel layout problems, they may be extended to ten or more; or they may be synthesized further into only three or four generalizations. In any case, these principles must be implemented further with relevant quantitative data before they are of practical value to the design engineer. Research in this area must be extended until every recommendation that is made for panel design can be supported by quantitative specifications of panel layouts. As recent experiments have shown, for example, it is no longer of any great value to point out that controls must be within easy reach of the operator. We must be able to indicate the limiting dimensions of panel layouts that will be satisfactory for known populations of operators under various conditions of operation. When data of this sort are available for all the principles of panel design, we will be in a much better position to evaluate equipment in current use and in the design stage. Then we can make practical, concrete recommendations for redesign that will satisfy the limiting requirements of human capacity.

Practical Compromises in Panel Design

Since panel layout problems have to be faced with the data at hand, however, it is well, at this point, to clarify some of the problems involved in the application of research knowledge to practical panel design. A considerable amount of compromise will be involved in any satisfactory panel layout. To begin with, it should be recognized that at least three major compromises will have to be made. In the first place, it should be apparent from the above discussion that it is not possible to arrange controls and displays in accordance with each principle of panel layout simultaneously. At our present stage of knowledge, satisfying the requirements of one principle will very likely mean violating those of another. For example, if controls are to be arranged in functional groups, it may not be possible to place all of them in positions for optimal use. Or conversely, if the position of a control is fixed by the fact that it requires the application of maximum force for operation, it may not be possible to include it in a functional group that must be located elsewhere on the panel. In order to effect such compromises most efficiently in a practical panel design problem, we should have some way of assigning priorities to each principle. For example, if long vigils are required in the operation of one type of equipment, perhaps the position of the operator should be considered first. Where speed and accuracy in the use of a limited number of controls is of prime importance, their location should be given first priority. Or where an operator has to handle a complex array of controls and displays, functional grouping and coding might be given first consideration.

Ideally any departure from the requirements of any principle should be based on experimental knowledge of what alternative arrangements and designs are least detrimental to performance. Since the existing literature does not afford adequate data on the efficiency of performance over a wide range of conditions, it will be necessary, in any practical situation, to proceed by empirical trial and error and try out various deviations from ideal design until the most satisfactory arrangement is found. But it should be pointed out that guesswork can eventually be eliminated when we have complete functions, describing the efficiency of performance over a wide range of variations in panel design. If problems this complex are to be solved scientifically, it is imperative that future research on human factors in panel layout be carried beyond the point of merely determining optimal arrangements and designs. We must be able to know beforehand how efficiency varies as a function of different amounts of departure from the optimum in any direction.

Perhaps a more serious type of compromise must be made between the ideal design for each individual control and display and the kind of design that might be called for when that item is to be placed in a particular context in a complex panel. It may turn out to be more efficient to violate some of the principles of the design of individual controls and displays in order to achieve more satisfactory layouts. Or the converse may be the case. For example, it has been pointed out that while certain control knobs or handles might be optimal for best operation of those controls, they may not be distinctive enough to prevent confusion of one control with another. The ideal would be to find knob designs that are both distinctive and easy to use, but in actual practice a compromise between these two ideals may have to be reached. Similarly, while it is desirable to have dials and indicators some optimal size, the demands of space and need for simplicity might be so great that it would be best from the operator's point of view to make them suboptimal in size. Here again, it is necessary to know what kinds of suboptimal designs are least detrimental to overall efficiency of operation in order to accomplish such considered violations of ideals for individual control and display design. For practical panel layout, therefore, future research on control and display designs cannot stop at determining just the optimal design but must afford us a complete function in every case, describing what varying degrees of deviation from the ideal will mean for operator efficiency.

Finally, it is necessary to make compromises between the panel layouts that would be ideal from the operator's point of view and what the design engineer can do. Not every rule dictated by studies of human capacities can be followed in the practical engineering situation. The engineer is limited by the structural features, mechanical and electronic performance, and the serviceability of the equipment. It is only within this framework that he can satisfy the requirements of the human factor. For example, sometimes he is limited as to where he can place a particular display because of the size of the power unit needed to drive it. Or, when the demands on space are too great, certain controls and displays will have to be placed beyond the practical working limits of the operator or assigned to a second operator such as a co-pilot or flight engineer. Whatever the case, specific engineering limitations must be accepted in realistic panel design. Then, within these limitations, the most efficient panel layout can be worked out. Of course, with further engineering advances, these limitations may be reduced and some of the problems facing the students of the human factor in panel layout may be obviated. For example, if satisfactory booster systems can be devised to help in the operation of controls, it may not be necessary to place a lever in a position for maximum pull.

The Evaluation of Panel Layouts

A second point that ought to be kept uppermost in mind in applying our knowledge of human capacities to panel design is that the ultimate criterion is the efficiency of operator's performance. Many factors which determine the operator's efficiency cannot be separated and studied in isolation. But it is still possible to do experiments by introducing changes in panel design that are indicated by the study of human capacities and objectively testing the performance of the operator. The example of redesigning the soldering operation is a good illustration of this point. Several aspects in the arrangement of the work situation were varied at one time, but the only criterion of success or failure of the changes was the output of the operators. Methods of evaluating operator performance in complex military tasks may not be as easy to devise as they are in the case of simple, repetitive, assembly-line jobs, but several techniques, described below, have been successfully carried over from industry (**7, 8, 20, 28**).

Equipment in the Design Stage

The best opportunity to apply our knowledge of human factors in panel design comes when the equipment is in the design stage. A variety of procedures for incorporating experimental data and principles into panel design have been mentioned in the literature (**7, 20, 24, 34**), but they may be summarized and synthesized in a few brief points. At the blueprint stage, a number of obvious points can be checked: the purpose of the equipment, the type and function of each control and display, the overall dimensions of the panels, the distance of controls and displays from each other and from the operator, the grouping of controls and displays on the panel. It is a simple matter to determine whether any of the known principles of panel design are violated, and corrections are easily made at this stage of design. Once all that is possible has been done with the blueprint, the next step is to construct mock-ups of the equipment to answer further questions that are bound to come up about alternative types of design. For example, in the design of aircraft cockpits, it is now possible to use a rather elaborate mock-up arrangement consisting of the Universal Test Seat developed at the University of Michigan and the Beindorf Fixed Preflight Trainer, which has typical aircraft controls and displays (**32**). The test seat permits wide variation in the position of the operator, and the trainer affords a realistic cockpit atmosphere in which the arrangement of controls and displays may be varied. Actually, simpler mock-ups are satisfactory for many purposes. For example, in studies of radar, simulated combat information can be presented on a scope and the speed and accuracy of subjects operating experimental control panels can be determined. Basically, all that is needed in a mock-up is a dummy panel of proper dimensions in which the controls and displays relevant to the layout problems of the moment may be manipulated for experimental test.

The final check-out of equipment will have to be made on experimental models of the equipment. These models should be flexible enough to permit the experimental introduction of alternative designs and arrangements of controls and displays without

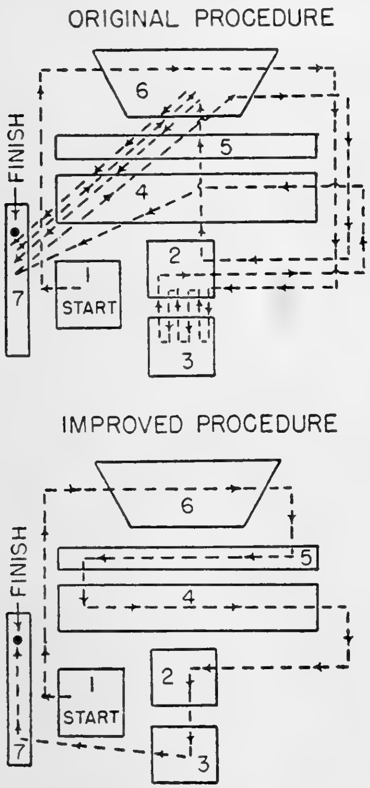

Fig. 10. Modification of the procedure of cockpit check-off to reduce waste motion. (From Channell, 7)

detracting from verisimilitude. At this stage, however, the process of evaluating performance becomes complex and requires fairly refined techniques such as those used in industrial time and motion studies. In the first place, we are dealing with complex behavior and have to resort to technical devices for recording it. Two major techniques that have been used in industry seem appropriate here: (1) micromotion analysis in which motion pictures of operators at work are used to indicate the motion paths that are followed in the use of controls, the amount of time spent in the operation of each control, the speed of reactions, and the relative use of parts of the body, and (2) eye movement records which indicate the frequency and duration of fixations of displays and the patterns of eye movements from one display to another. In addition, wherever possible, it is of value to have an overall measure of the speed and accuracy of performance such as the rate at which information can be put through equipment used in communication centers, the accuracy or range and bearing readings in radar operations, the length of time a man can work efficiently at a plotting board under standardized conditions. Finally, by appropriate questionnaire techniques the opinion of typical operators should be obtained on such questions as the comfort of the seating arrangement, the importance of controls and displays, and their preference for alternative arrangements of the panel. In all of these measures of performance and opinion, it is important to use a reasonably large sample of subjects, representative of the operators who are likely to use the equipment in respect to age, education, intelligence, physical size, and aptitude.

Equipment in Current Use

In actual practice, much of this evaluation work has to be done on equipment in current use where there is little opportunity to manipulate panel layouts or introduce changes in design. The problem at this stage is primarily one of determining whether or not the panel design and arrangement embodies the best of present day knowledge and principles of the layout of controls and displays. For different types of equipment different problems will have to be faced, of course. Some control and display panels are difficult to use because they are so complex; others are relatively simple but require a person to operate a small number of controls with great precision or to work for long hours. In any case, the same methods of evaluation may be used. Experienced operators, representing a fair sample of the population that uses the equipment, should be studied by means of the time and motion techniques of micromotion analysis, eye movement recording, opinion survey, etc. Up to the present time, these procedures have been applied with some success to the evaluation of aircraft cockpits (**7**) and radar consoles (**24, 34**) but their use should be extended to other types of equipment.

Usually only minor changes can be introduced in current equipment without getting into the problem of complete redesign. But wherever possible, they should be made, and then the equipment should be re-evaluated for efficiency of operation. In cases where changes are not feasible, it may be possible to recommend changes in the training of operators or changes in the procedures they follow in working the equipment. For example, one study has shown that the cockpit of the DC 4 is poorly arranged for the existing procedures of checking out the instruments and controls before take-off (**7**). It was pointed out that this difficulty could be rectified by changing the sequence in which the pilot makes his checks (Fig. 10). In another case, it was suggested that the accuracy of reaching for controls in the cockpit could be improved without redesign of equipment by training pilots in "blind reaching."

Conclusion

From the foregoing discussion, it is apparent that more research is needed on problems of panel layout. At the present time, a good start has been made toward specifying human requirements in the design of panels. General principles and some quantitative data are available on six major aspects of panel layout: (1) the practical limits of the working area, including the position of the operator and the orientation of the panel, (2) the optimal location of controls and displays, (3) the importance and frequency of use of controls and displays, (4) the patterning of controls and displays, (5) the distribution of work over the body parts, and (6) the confusion of controls and of displays. It should be remembered, however, that the research to date represents only a start in the solution of practical problems. Psychological experiments must be able to give the design engineer complete quantitative data on human performance as a function of all the important factors in panel layout. Furthermore, since it is unlikely that we will be able to design equipment panels that incorporate the optimal specifications for all factors at once, it is of practical importance to know, in each case, how performance varies over a wide range of conditions deviating from some optimal design or arrangement. Only then will it be possible to design equipment in keeping with the self-evident principle that performance is a joint function of all the factors in panel layout.

Of equal importance with the study of human capacity is the investigation of the problems involved in the practical application of research knowledge to panel design problems. The basic problem here is the development of techniques for measuring and evaluating human performance in complex man-machine situations. Some headway has been made in adapting the time and motion techniques of industry for measuring performance on certain kinds of equipment. But this work must be extended to other types of equipment, and new techniques must be developed if all the major aspects of human performance are to be assessed. The importance of developing valid criteria and methods for evaluating the suitability of panel layouts in terms of the efficiency of human performance cannot be overemphasized. Without adequate evaluation techniques, the successful application of research knowledge to practical equipment problems is virtually impossible.

References

1. Barnes, R. M. *Motion and time study.* New York: John Wiley, 1940.

2. Barnes, R. M. *Work methods manual.* New York: John Wiley, 1944.
3. Bartlett, F. C. *Instrument controls and displays.* Univ. Cambridge Psychol. Lab., Md. Res. Counc., Roy Naval Personnel Res. Committee, Report No. 43/81, 10 December 1943. (RESTRICTED)
4. Bartlett, F. C. *Instrument controls and displays—efficient human manipulation.* Great Britain: Flying Personnel Research Committee. Report No. 565, December 1943.
5. Brown, C. W., Ghiselli, E. E., Jarrett, R. F., & Minium, E. W. *Speed and accuracy of reaching for controls in different areas from the prone position.* Aero Medical Laboratory, Report No. MCREXD-694-4H, 28 January 1948.
6. Browne, R. C. *Suggestions for improving beam approach indication methods.* Great Britain, Air Ministry, FPRC 418 (b), July 1943. (RESTRICTED)
7. Channell, R. C. *An analysis of pilot's performance in multi-engine aircraft (R5D)* Special Devices Center, Office of Naval Research, 1947. Report on Project 20-0-1, Psychological Corporation.
8. Channell, R. C., & Tolcott, M. A. *The use of human engineering data in equipment design problems.* Special Devices Center, Office of Naval Research, 1948. Report on Project 20-F-2, Psychological Corporation.
9. Chapanis, A., Garner, W. R., Morgan, C. T., & Sanford, F. H. *Lectures on men and machines.* Special Devices Center, Office of Naval Research, 1947. Report No. 166-I-19. (RESTRICTED)
10. Clark, W. G., Henry, J. P., Greeley, P. O., & Frury, D. R. *Studies on flying in the prone position.* U. S. NRC-CAM Contract No. OEMcmr-288, CAM Report No. 466, August 1945. (RESTRICTED)
11. Fitts, P. M. Psychological requirements in aviation equipment design. *J. Aviat. Med.*, 1946, **17**, 270–275.
12. Fitts, P. M. A study of location discrimination ability. In: Fitts, P. M. (Ed.) *Psychological research on equipment design.* U. S. Government Printing Office, 1947, 207–217.
13. Fitts, P. M., & Jones, R. E. *Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls.* Aero Medical Laboratory, Army Air Forces, Report No. TSEAA-694-12, 1947.
14. Fitts, P. M., & Jones, R. E. *Psychological aspects of instrument display.* I. *Analysis of 270 "pilot-error" experiences in reading and interpreting instruments.* Aero Medical Laboratory, Army Air Forces, Report No. TSEAA-694-12A, 1947.
15. Foley, F. M. *Current philosophy of instrumentation in U. S. Navy airplanes.* U. S. Navy, June 1943. (Abstract)
16. Forbes, T. W., Garner, W. R., & Howard, J. G. *Flying by auditory reference ("Flybar").* Psycho-Acoustic Laboratory, Harvard University, OSRD Report No. 5123. 1 June 1945. (RESTRICTED)
17. Grether, W. F. *A study of several design factors influencing pilot efficiency in the operation of controls.* Aero Medical Laboratory, Army Air Forces, Report No. TSEAA-694-9, 1946.
18. Hugh-Jones, P. *Some physiological aspects of tank driving controls. Part I. The maximum push exertable on a foot pedal and how this varies with the position of the pedal relative to the seat.* Great Britain, MRC-MPRC Subcommitte on AFU, BPC 44/341. P. A. 137, May 1944. (RESTRICTED)
19. Hugh-Jones, P. *Some physiological aspects of tank driving controls. Part III. The maximum force exertable on hand controls and how this varies with different positions of the control relative to the seat.* Great Britain, MRC-MPRC Subcommittee on AFU, BPC 45/410. P. L. 158, January 1945. (RESTRICTED)
20. Kelley, G. A. *Notes on aircraft instrument panel design.* Unpublished notes, November 1945. (Abstract)
21. King, B. G., Morrow, D. J., & Vollmer, E. P. *Cockpit studies—the boundaries of the maximum working area for the operation of manual controls.* Naval Medical Research Institute, National Naval Medical Center, Bethesda, Maryland, Project X-654, Report No. 3, July 1947.
22. Kleem, O., & Sander, F. Experimentelle Untersuchungen über die Form des Handgriffes an Drehkurbeln. *Prakt. Psychol.*, 1923, **4**, 300–302. (Abstract)
23. Leyzorek, M. *Mounting angle of a VJ remote radar indicator and its effect on operator performance.* Special Devices Center, Office of Naval Research, Report No. 166-I-41, February 1948. (RESTRICTED)
24. Lipschultz, H. L., & Sandberg, K. O. W. *Maximum limits of working areas on vertical surfaces.* Special Devices Center, Office of Naval Research, Report No. 166-I-8, 1947.
25. McFarland, R. A. *Human factors in air transport design.* New York: McGraw Hill, 1946.
26. McGehee, W. *Comparative study of pilot fatigue resulting from extended instrument flight*

using the standard AAF and British instrument panels. U. S. Navy, Bureau of Medicine and Surgery, Final Report, Project TED No. ATL-R601. (No date) (Abstract)

27. Mogenson, A. H. *Common sense applied to motion and time study.* New York: McGraw Hill, 1932.
28. Orlansky, J. *The human factor in the design of stick and rudder controls for aircraft.* Special Devices Center, Office of Naval Research, 1948. Report on Project 20-M-1c.
29. Patt, D. I. *Cockpit dimensions in relation to human body size.* Engineering Division, Aero Medical Laboratory, Army Air Forces, Report No. TSEAL-3-695-32TT, April 1945.
30. Patt, D. I., & Randall, F. E. *Principles of seating in fighter type aircraft.* Engineering Division, Aero Medical Laboratory, Army Air Forces, Report No. TSEAL-3-695-58, September 1945.
31. Randall, F. E., Damon, A., Benton, R. S., & Patt, D. I. *Human body size in military aircraft and personal equipment.* Air Materiel Command, Army Air Forces, Tech. Report No. 5501, 1946.
32. Randall, F. E., & Patt, D. I. *Universal test seat and cockpit mock-up.* Engineering Division, Aero Medical Laboratory, Army Air Forces, Report No. TSEAL-3-695-32RR, March 1945.
33. Rubarth, B. Untersuchung zur Bestgestaltung von Handheften für Schraubenziehen und ähnliche Werkzeuge. *Industr. Psychotech.*, 1938, **5**, 129–142. (Abstract)
34. Sandberg, K. O. W., & Lipschultz, H. L. *A survey of the importance and use of controls and displays on radar console panels (a contribution to panel layout).* Special Devices Center, Office of Naval Research, Report No. 166-I-17, 1947. (RESTRICTED)
35. Sandberg, K. O. W., & Lipschultz, H. L. *Relative performance for cranking a handwheel at different positions on a vertical surface.* Special Devices Center, Office of Naval Research, Report No. 166-I-22, 1948.
36. Weitz, J. The coding of airplane control knobs. In: Fitts, P. M. (Ed.) *Psychological research on equipment design.* U. S. Government Printing Office, 1947, 187–198.

CHAPTER 7 "Arrangement of Equipment" pages 177-198, have been cut out of this book and upgraded to Confidential

PART III.

AUDITORY PROBLEMS

Chapter 8

AUDITORY SIGNALS

W. R. GARNER

The Johns Hopkins University

Introduction

We can consider, for our purposes, that a display is any means of presenting information. An auditory display, then, is any method of presenting information through our sense of hearing. While speech is by far the most common form of auditory display, it is not the only one, and in this section we are interested in those auditory displays that do not require speech. The chime of a clock, the whistle of a factory, a fire siren, a police whistle, all are auditory displays. Likewise, any specially designed set of auditory stimuli which transmit information of a particular type are auditory displays.

Use of Auditory Signals

With the single exception of Sonar, auditory signals have not been widely used on submarines. Lack of use, however, does not necessarily mean that auditory signals would prove valueless if used correctly, and in situations where they would be most efficient. Some parts of a submarine are practically nothing but display areas. Information of all kinds must be read from meters, special visual displays, radars, and dials, and this information must frequently be used by people who cannot look at the information directly. Thus the information must be transmitted, either by telephone or by some special remote visual display. Information must be transmitted, for example, to torpedo men, to radar operators, and to gunnery men.

Auditory signals could be used profitably in many of these situations. Whether it is worth while to develop an auditory system, or whether one will be feasible at all, depends on the particular circumstances, and on the possible advantages of the auditory display. There are essentially three potential advantages in the use of auditory displays, and all three advantages might be realized to some extent in submarine operation.

Speed of Communication. One of the chief advantages of an auditory signaling system appears in situations where only a limited amount of information is required. The auditory signal can be designed to give just the kind of information required, and thus can give that information in greater quantity and faster. The *radio range*, for example, is an auditory signal which provides a pilot with only one kind of information—the azimuth position of the plane. The information, however, is presented to the pilot almost continuously; considering the speed with which the pilot can use the information it is indeed continuous.

Substitute for Visual Display. In a sense, then, simplified auditory signals can be used as a substitute for speech when only a limited kind of information is needed. Auditory displays are also useful wherever the total amount of information presented to an operator visually is too great. The aircraft is again a good example. The pilot must look at a multitude of visual displays, and it is frequently desirable to present some of that information by ear to relieve the load on his eyes.

Automatic Transmission of Information. A third type of situation in which auditory signals can be advantageous occurs whenever it is necessary, for some reason or other, to transmit information automatically.

There are two main reasons for automatic transmission. The first of these reasons is that automatic transmission does not require the use of human operators on the input side of the communication system. Whenever there is a shortage of manpower, or a shortage of space, it is desirable to avoid the use of human operators.

The second reason for using automatic transmission of information is that faster communication is possible. Whenever a human operator is brought into the picture, we have the inevitable delays due to his relatively slow reaction time, plus the fact that the human being has to comprehend a situation before he can transmit information about it. An excellent example of this is gun control on ships. Before a man can tell a gun director operator where to point to pick up a target, he has to determine that there is a target, then determine its position in two or three coordinates, and then transmit all that information by telephone. An automatic signal generator could produce the appropriate signals just as fast as a human operator could place some pointer, electronic or mechanical, on the target display. The total savings in time here could easily approach ten seconds or more. Attention should also be called to the possibility of allowing dials and indicators to do their own "talking." An automatic annunciator was developed during the war (**6, 7**) which proved highly successful in laboratory tests, but which has not yet been put to practical use. This automatic transmitter used a battery of magnetic tapes on which were recorded various speech signals, such as numbers or words. A relay selector made it possible to convert the correct reading on a dial into words or numbers selected from the appropriate magnetic tapes, and this signal then was transmitted automatically to the listener. While this device was developed principally to provide automatic transmission of a limited range of speech signals, it could be just as easily used to provide a variety of special tonal signals. In effect, this device makes it possible for a dial to be read automatically, and for the dial reading then to be transmitted without a human operator ever seeing the dial or talking over a telephone network. Its possible uses are many, and an investigation of its potentialities should be undertaken by the Submarine Service.

Kinds of Information

There are different kinds, or degrees of precision, of information which we might want to transmit in a given situation, and the utility of a given type of auditory signal depends a great deal on the kind of information required.

Yes-No Information. The simplest kind of information is of the yes-no, either-or, type. Did an event happen or didn't it? Should an action occur, or shouldn't it? For example, a gunner can be told to fire a gun in a single positive order. If the gun should not be fired, no signal is needed. The tonal signal in this case can itself be a yes-no type of signal. If no action is to occur, no signal occurs. If the action is to occur, the signal is heard. If no other information is to be transmitted with this particular communications system, then the kind or quality of the tone is rather incidental. The only real problem is whether the signal can be heard, since there is no possibility of confusing signals. Of course, if a particular communications system has to provide information about several possible actions, then we need as many easily distinguishable tones as we have actions, but no more.

Directional Information. A second level of information can be called directional, or qualitative. Here we need not merely audibility of a signal, but directionality, as well. For example, we might want a tonal signal which indicates limits of tolerance, but which also indicates the direction of deviation from those tolerance levels. Is a gun pointed too high or too low? Is the rate of fire too fast or too slow? In the psychological sense, yes-no information requires only audibility of a signal. Directional information requires differential sensivity of the greater-than or less-than type.

Quantitative Information. A still more precise kind of information requires not only identification of direction, but also of amount. Here, of course, we get into all the problems of psychological scaling. If we want to provide automatic transmission of information to a fire-control radar operator, then we must be able to tell the operator exactly where to position his director. It is not enough to provide simple directional information—at least, if we want a reasonably fast response on his part. Simple directional information can be used if the operator moves the director in the stated direction at a constant speed until the signal tells him that he is now "on." Lack of quantitative information, however, makes it impossible for the director operator to adjust his speed of movement to the amount by which he is off the target, so that he can move rapidly for large misadjustments, and slowly for small ones.

We could, of course, have combinations of various kinds of information in one tonal signaling system. The more we combine different kinds of information into one auditory system, the greater the psychological problems become. In fact, one of the most important psychological problems in this area is the determination of the usefulness of tonal signals. How many different kinds of information can be provided with tones? What characteristics of tones make it easy for them to be ignored, and what characteristics make one signal stand out against other signals? We know very little about the answers to questions such as these.

In considering, or designing, a tonal signaling system, it is important to know precisely what kind of information is required. A good general principle, not only in auditory signals, but in all displays, is never to provide more information than is needed or can be used. Too much information only confuses the listener, and slows him down in using the signaling system. Also, the more information provided, the more difficult are the discriminations required.

Illustrations of Auditory Signals

Auditory signals, other than speech, can be classed roughly into two groups. The first of these we will call occasional signals. They are signals for which no special communications system is set up, but which occur for anybody to hear who happens to be present. These signals do not provide at all a continuous source of information, for they occur only on some specific event. The alarm of a clock, the siren of the fire engine, the factory whistle, are all occasional signals. This type of signal is not much of a problem, since about the only requisite is audibility. Problems of discriminability rarely occur.

The other class of signals involves a specific communication system which provides a continuous source of information, although not necessarily a continuous transmission of information. Such signals are less of a display, in the sense that they do not provide information which may be used by anyone who happens to hear them. The continuous signals are usually there because it is expected that some action will occur as a result of them. These signals are specific to a particular series or group of related actions.

There are not many continuous signals in use, although it is quite probable that they could be used much more than they actually are at present. We shall discuss some of the work done on two types of auditory signal—the *radio range* and *Flybar*—as illustrations of research which is specific to problems of auditory signaling. These two types of signals are, of course, not used on submarines. However, if research is undertaken on the possibility of using tonal signals on submarines, the program of exploration and of testing would presumably follow the lines of the research programs undertaken with these two signals.

Radio Range Signals

Radio range signals provide aircraft pilots with directional flight information. The signals are produced by two directional radio beams with their axes crossed at right angles. The indicated direction for the pilot then lies along the path where the field strengths of the two beams are equal. One of the two beams transmits the letter A in telegraphic code, and the other the letter N. These two signals are interlocked in such a way that the pilot hears a continuous tone if the two signals have equal intensity, but an A or an N if he flies too far to one side or the other.

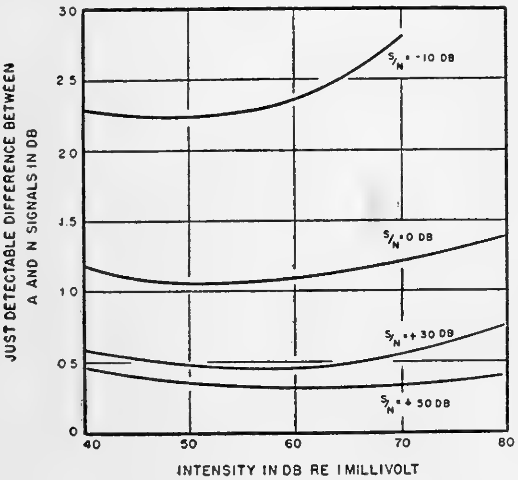

Fig. 1. The effect of intensity and signal-to noise ratio on the discrimination between A and N signals in the radio range. (After Flynn *et al.*,**5**)

From the point of view of the psychological problem involved, the radio range is a very simple type of auditory signal. It provides directional information of only one kind, and the psychological problem is one of simple discrimination. If the pilot can make the auditory discrimination, he can fly in the correct path; if he cannot make the discrimination, his flight will be incorrect or erratic.

A series of investigations was reported by Flynn *et al.* (**5**) shortly after the war. These experiments were concerned with the factors which affect the pilot's ability to discriminate the radio range signals. We shall discuss some of this research as an illustration of the kinds of problems which must be met.

The Effect of Intensity

One experiment was designed to determine the effect of signal intensity on the discrimination, when the discrimination is made with a fairly high noise background. Fig. 1 shows some of the results obtained. Each curve is for a different signal-to-noise ratio, expressed in decibels. It can be seen that there is an optimal intensity for each signal-to-noise ratio, but that the S/N ratio itself has much more effect on the discrimination than overall intensity.

This kind of result is not unlike that found with many more academic experiments on the intensity discrimination of tones. When research is specifically related to the use or operation of some instrument, however, the results frequently seem more practical. In this case, for example, the fact that there is an optimal intensity for each signal-to-noise ratio provides immediately useful information about the operation of radio receivers.

Frequency Characteristics of the System

Band-pass Filters. Another experiment was concerned with the effect of noise filters on the discrimination of the radio range signals. Here discrimination was compared for the normal case (no filtering) with the use of a narrow-band-pass filter—a filter which rejected all noise frequencies above and below a narrow band around the audio frequency of the signal. Fig. 2 shows that such filtering improves discrimination between the A and the N signals considerably. This experiment again was done with a practical point in mind, but the results and the experiment itself have general applicability.

Type of Earphones. Still another experiment was performed to test the practical conclusion derived from the experiment just cited. Some earphones have an essentially flat response over a wide range of audio frequencies; i.e., the acoustic energy is the same for any frequency if the voltage across the earphone is the same. Other phones tend to produce much more energy at one frequency than at others: they have a peaked response. If the resonance peak of the earphones corresponds to the signal frequency, then the earphone is essentially a band-pass filter, and discrimination should be better with the peaked earphones than with the flat earphones. The experimental tests showed that discrimination actually was better with the peaked phones, although the improvement was not as much as when the narrow-band-pass filter was used. The practical conclusion of this experiment was, then, that distortion-producing phones are better for this particular type of signal than flat-response phones.

Special Instruments

Expanders. Other experiments were even more specific, in the sense that they were only tests of particular pieces of equipment which might be used with radio range instruments. One of these special gadgets was a signal expander—a device which amplifies large signals more than it amplifies small signals. Thus it should make differences in signal strength more different, and improve discrimination. The tests showed that discrimination actually was improved, and demonstrated the feasibility of the particular instrument.

Pitch Modulators. Another device tested was one which translated intensity differences into frequency differences. The tests showed that pitch discrimination was better than intensity discrimination at favorable S/N ratios, but was not as good when the signal was appreciably less intense than the noise.

Summary of Radio Range Experiments

The experiments on the radio range signal are illustrations of the kinds of research done on the simplest type of auditory signal. Only simple discrimination is required, and much of the research was concerned solely with factors which affect the intensity discrimination of pure tones. This type of research is fundamental to our understanding the particular signal under investigation, as well as to our understanding the basic operation of the ear. Other experiments were concerned solely with the specific means of producing the signal, and can be considered applied research or tests. Such experiments, on the other hand, contribute little to our understanding of the basic phenomena of hearing, and the experiments would in fact be quite unnecessary if we knew enough about how the ear works.

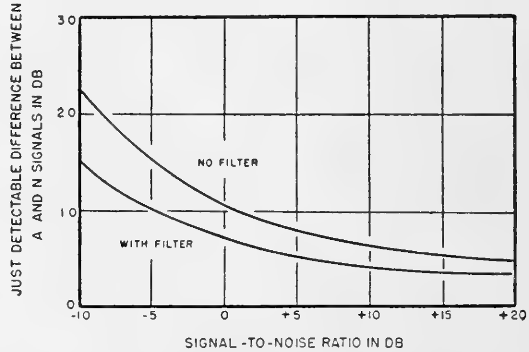

Fig. 2. The effect of signal-to-noise ratio on the discrimination between the A and N signals of a radio range with and without a narrow band-pass filter. (After Flynn *et al.*, **5**)

On submarines, simple signals like the radio range could be used to provide lateral directional information for positioning guns or radar equipments. Continuous and automatically transmitted information could increase both speed and precision of gun positioning, for example, and would certainly make manpower shortages less critical.

Flybar

The word *Flybar* is an abbreviation of "flying by auditory reference" and has been applied to attempts to provide auditory signals for use in flight. In 1936 de Florez (**1**) demonstrated that pilots can fly by means of auditory signals presented in earphones, even though his method did not involve a fully developed system of auditory signals. During the late war more extensive research was undertaken, and the results are presented in a report by Forbes *et al.* (**7**). A shorter report of certain of the studies was later made by Forbes (**6**). The account given here concerns these latter experiments.

The Problem

The chief reason for wanting auditory signals in flying is to relieve the load on the pilot's eyes. Present-day aircraft have a multitude of dials and instruments which the pilot must watch all the time, and it was felt that it would be a great advantage if some of this load could be shifted to his ears.

A pilot can keep his plane flying along a predetermined path if he has three kinds of information—information about the three dimensions in which the plane is flying. He needs information about the turn of the plane, its bank, and its tilt. In actual practice, airspeed is equivalent to the dive-climb (tilt) indication, since airspeed increases if the nose goes down, and decreases if the nose goes up. The problem, then, was to develop auditory signals which would transmit information about the three basic indications. In the actual signals, airspeed was almost always used instead of a dive-climb indication.

An auditory signaling system of this kind is extremely complex compared to the radio range. Here information of a quantitative nature must be provided about three different kinds of action simultaneously. To provide simple yes-no or directional information in this situation would not be nearly precise enough to enable the pilot to fly with certainty.

Experimental Developments

Preliminary Experiments. Before any full-scale signaling systems were set up, preliminary experiments were undertaken to explore various possibilities for signals. These showed that intensity differences between the two ears gave a very inexact indication of direction, and that phase differences were likewise poor. More will be said of these problems later. As a result of the preliminary tests, however, several complete systems were later tried in an airplane pursuitmeter task.

Three-Tone Signals. One set of signals made use of three different tones in each ear, plus an additional tone to indicate bank. One of the three tones had a low frequency, and another had a high frequency. The third tone varied in frequency between the other two, and the pitch of this middle tone indicated airspeed. When the middle tone approached either the higher or the lower frequency, beats could be heard, which set airspeed tolerance limits. To indicate a turn, the tones were interrupted in the ear toward which the turn was occurring, and the faster the rate of interruption, the faster the rate of turn. Bank information was produced by the insertion of a low-frequency tone on the side of the low wing. Three different frequencies indicated three degrees of bank.

These signals were not very successful, primarily because the subjects in the experiment tended to listen to one or another of the three indications to the exclusion of the other two signals. As a result, one indicawould frequently go far out of bounds, while one or two of the others would remain well centered.

Sweeping-Tone Signals. Other combinations of signals were also tried out, but the one which was most successful combined all three indications into a single tone, which was made to increase in intensity in one ear while it decreased in the other ear. An illusion of movement was thus produced to indicate turn. The direction of movement indicated the direction of turn, and the rate at which the movement was repeated indicated the amount of turn.

While the tone was changing in intensity, it was also changing in frequency. The bank of the plane was indicated by the relative frequency of the tone as it appeared to move from one ear to the other. If the frequency was high when the tone seemed to be in the left ear, and low when in the right ear, then the left wing was high. The opposite relation indicated that the right wing was high. The amount of the bank depended on the amount of frequency change as the tone moved from left to right. Airspeed indication was provided by introducing a short, fast "beep" into the tone. This beep was actually a rapid frequency modulation, and the rate of the beep depended on airspeed. For each sweep of the tone from ear to ear, the time was divided into two parts. During the first part, a standard beep rate was heard, and the second part gave a beep rate which depended on airspeed. The use of a standard rate had been shown necessary in some of the earlier attempts which we have not described here. This type of signal was found to be reasonably successful.

Link Trainer Tests. Later tests on a Link trainer showed that pilots could fly a fairly straight course with these signals. The tests also showed that pilots could learn to fly the trainer with the auditory signals as fast as with the visual indications. The described signals were not necessarily the best possible signals, but they were good enough to demonstrate that the use of auditory signals was entirely feasible.

Conclusions from Flybar Experiments

The Flybar experiments were not designed to provide basic knowledge about the use of auditory signals. They were the outgrowth of an attempt actually to develop a set of signals for use in aircraft. Nevertheless they provided information about some of the basic problems and requisities of such signals. The following might be considered the basic considerations for a successful system of auditory signaling, in submarines or elsewhere.

1. The signal should sound real; i.e., it should sound as nearly as possible like what is actually happening. This is particularly true if the signal indicates motion or position; it ought to sound like motion or position. In other words, the signals should not depend on a remote symbolism. Turn could be indicated by the intensity of a monaural tone, but such indication would be very poor because it does not sound like what is actually happening.

2. Whenever more than one kind of information must be presented to an observer simultaneously, the signals should be designed to prevent listening to only one signal at a time. In the Flybar experiments, this was accomplished by combining all indications into one basic tone.

3. Whenever directional or quantitative information is to be indicated, some standard signal must be provided to which the listener compares the critical signal. Absolute psychological standards are very poor, especially in hearing. For example, in some of the signals, airspeed was indicated by a "put-put" which varied in rate much as a motor sounds when it changes speed. But this signal was usable only when some standard "put-put" rate was provided. The observers could not remember what "normal" speed was. Nor can they remember what a "normal" intensity or frequency is. However, in spite of the fact that the absolute pitch of a tone, or its intensity, cannot be identified, it is an easy matter to tell that one tone is louder than another, or that it is higher in pitch. Thus, problems of directional and quantitative information all become matters of auditory discrimination in the usual academic sense of discriminating one tone from another.

Although the Flybar experiments demonstrated that auditory signals are feasible in a communication system calling for three kinds of information, it is questionable whether the signals actually used were the best signals. It would be very helpful to have a program of further developmental work on complex signals. If such signals can be used on aircraft, there is little reason why they cannot also be used successfully on submarines. We will not know the extent of their possible uses until more practical development work has been done.

Basic Psychological Problems

There are two ways of approaching the problems of auditory signaling. The first way is to design various sets of signals and try them out. It is a cut-and-try method, and is essentially the way the Flybar signals were developed. We do not know enough about the psychology of hearing to determine beforehand which signals will work and which will not. It is true that we have enough information so that we do not have to start completely from scratch, but a great deal of trial-and-error is still necessary. With a little luck, such a procedure might turn up very useful signals, and might be the fastest way of providing signals for any given kind of information system.

The second approach is to do basic research in those problems of hearing which are most relevant to the design of auditory signals. If we had sufficient information about the psychology of hearing, it would be possible to design a set of auditory signals with considerably less guesswork. The cut-and-try procedures could be shortened or even eliminated. Basic research relevant to problems of auditory signaling has the additional advantage of providing data that are not related exclusively to a particular informational system. The data would be useful in the design of auditory signals for any and all purposes. In the long run, it is probable that time spent on basic research would be of more advantage than time spent on cut-and-try designs of specific signalling systems.

In the rest of this section, we shall discuss some of the basic psychological problems which are relevant to the design and use of auditory signals. For example, suppose that a signal is required which will provide continuous information about one meter reading. If there is a "correct" reading, and action is required whenever the reading becomes too high or too low, then the listener must know when the reading becomes too high or low. He must, in other words, be able to discriminate between different tones in the signal. If, in addition, the listener must know how *much* too high or too low the reading is, then the auditory signal must have some quantitative aspect. Thus we get into problems of psychological scaling, which are problems of the quantitative dimension of tones.

Likewise, if we want to transmit lateral directional information—to a radar operator, for example—then we need to know the various ways in which a lateral localization of a tone source can be simulated. We then face the problems of binaural localization of tones. Basic research on this problem will make it possible to provide a lateral indication with little difficulty. In the Flybar experiments, as an illustration, a great deal of time was spent in attempting to find a good azimuth indication because it was not known at that time which types of signal will provide lateral displacement of an apparent tone source. If more basic information had been available, this time could have been saved.

These basic psychological problems, then, are the problems which will arise when an auditory signaling system is designed. They are the problems about which we need much more information if an auditory signal is to be designed with a minimum of guesswork.

Simple Discrimination

By simple discrimination we mean the ability to distinguish between two pure tones in either intensity or frequency. It is not often in real life that we are called upon to discriminate between two tones differing only slightly in intensity or frequency, and yet these basic functions provide vital information by the use of tones. Since, as we pointed out above, absolute psychological quantity is a very nebulous thing, most information must be transmitted by means of differences between tones. The listener must always have a standard tone which serves as his reference or base line. What are some of the factors that affect the discrimination between tones?

Intensity Discrimination

The classic research on the discrimination between intensities of tones was done by Riesz (**27**) in 1928. His experiments explored the effects of both intensity and frequency on intensity discrimination over fairly wide ranges. His work was comprehensive, and yet does not provide enough information. He used a technique of beats, in which the observer's task was to detect the presence of beats between two tones. It is quite likely that another method would not yield the same functions as he obtained.

Futhermore, many things besides intensity and frequency affect intensity discrimination. Montgomery (**24**) listed quite a few of the factors which affect the discrimination, although it is unlikely that one list would cover all possible effects. Garner and Miller (**14**) showed that the duration of a comparison tone affects intensity discrimination, but here again the technique was so limited that the complete story is not available.

Later experiments by Garner (**13**) showed that the effect of duration on intensity discrimination depends on how the tones are presented. If a short comparison tone is contrasted with a longer standard or reference tone, discrimination is much poorer than if both standard and comparison tones have a short duration. (See Fig. 3.) Unfortunately, the story does not quite end there, because in this same experiment it was shown that this statement is true only when the silent interval between the standard and the comparison tones was on the order of half a second. With very short silent periods, discrimination was as good with one method of presentation as with the other.

The problem of duration, incidentally, is one of the more important ones needing research for auditory signaling applications. In auditory signaling, the total amount of information presented in a given period of time depends directly on the speed with which any single item of information can be given. Thus it is necessary to present information as rapidly as possible, although it cannot be presented so fast that the discrimination process breaks down. We need to know, then, what the limits of discrimination are for short tones.

Other factors which undoubtedly have an effect on intensity discrimination, but about which we know very little, are the repetition rate of tones (how many discriminations per unit time), the transition between tones (whether abrupt or gradual), and knowledge about when tones are going to occur.

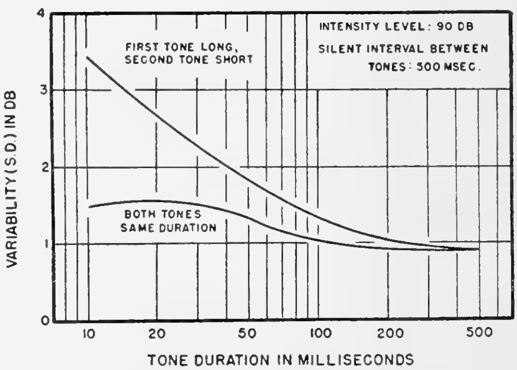

Fig. 3. The effect of tone duration on intensity discrimination

For the upper curve, the standard (first) tone was always 500 milliseconds, and the comparison tone had the duration indicated on the abscissa. For the lower curve, both tones had the duration indicated on the abscissa. With shorter silent intervals between tones, both curves resemble the upper one shown here. (After Garner, **13**)

Frequency Discrimination

In 1931 Shower and Biddulph (**28**) reported an extensive series of experiments on frequency discrimination. Their results show the effect of both intensity and frequency on frequency discrimination. They used a warble technique in which the tone was subjected to a small amount of frequency modulation. The observer's task was to detect the modulation. Here again, functions obtained with other procedures are needed because of the restricted nature of the method.

Turnbull (**35**) has shown how duration affects frequency discrimination. His technique provided a standard duration tone followed after a silent interval by a tone of shorter and variable duration. Discrimination became rapidly worse at durations of less than 100 milliseconds, and was almost impossible below about 20 milliseconds.

Experiments in progress by the author indicate that frequency discrimination as a function of duration depends on whether a short tone is compared to a long one, or whether two short tones are compared. Particularly at the very short durations, discrimination is better when both tones have the same duration. Doughty and Garner (**2**) have shown that most tones in the middle and higher range of audio frequencies must last for about 10 milliseconds for them to sound like a tone at all. Later experiments by these same authors (**3**) also showed that tones tend to sound lower in pitch when they become very short, but that this pitch loss is relatively slight and occurs only at very short durations (under 50 milliseconds).

Experiments by Harris (**16**) indicate that frequency discrimination is considerably poorer when a masking noise is great enough almost to mask the tones. When tones are about 10 to 15 db above the masked threshold, however, the presence of the masking noise has little effect.

Factors which affect intensity discrimination undoubtedly also affect frequency discrimination, and research is needed to determine just how much of an effect is present, and where the effects occur.

Complex Discrimination

An area where a great deal of research is needed is in the discrimination of complex tones of various sorts, and in the complex discrimination process. The literature in this area is very weak.

Complex Tones

We know practically nothing about simple intensity and frequency discrimination of complex tones. Any research in this area would be useful, but particularly we need to know how discrimination of complex tones compares with discrimination of pure tones. Is frequency discrimination better when a tone has a high harmonic content, as in square waves and pulse wave forms? It may be that discrimination can be improved considerably by using such tones. Miller (**23**) has studied intensity discrimination with noise, and found discrimination to be about the same as for tones.

Discrimination of Spectrum Changes

Another kind of problem about which we know practically nothing is the discrimination of spectrum changes. The recent research on speech seems to indicate that we understand speech by discriminating the changes in spectrum through time. We need to know discrimination limits for harmonic content. There is little question that humans can tell the difference between a pure tone and one of high harmonic content, but we do not know how precise that discrimination can be. A possible step in this direction has been taken by Karlin (**21**), who presented data on the discrimination of the pitch and loudness of noise spectra.

Glide Tones

Pepinsky (**26**) reports an experiment on the perception of frequency-modulated tones. His tones changed frequency in one direction over a short period of time, and he reports that the ability to detect the change in pitch depends both on the amount of the frequency change and on the duration of the change, a greater frequency change being required for shorter durations. His experiments, however, were done at only one frequency, and do not provide sufficient data for all purposes. More research is needed on the perception of this type of tone, and comparisons between frequency discrimination of glide tones and abruptly changing tones would be valuable.

Simultaneous Frequency and Intensity Discrimination

Another kind of experiment which is particularly pertinent to auditory signaling problems is the complex discrimination type of experiment. What happens to frequency discrimination when both tones are not of the same intensity, or when both frequency and intensity change simultaneously? Does discrimination deteriorate seriously when an observer must report changes in either intensity or frequency, or both simultaneously? Can an observer alternate intensity discrimination with frequency discrimination without one or the other breaking down? These questions and many similar ones cannot be answered without further research.

Pattern Discrimination

Still another area of research which would be most valuable for the problems of auditory signalling is that of auditory patterns and pattern discrimination. What little work has been done in this area has been concerned with time patterns, usually with respect to learning telegraphic code. Taylor (**34**) has summarized this research, although very little of it is directly useful for the design of auditory signals. Other types of problems concern intensity and frequency patterns. Can one frequency pattern of a series of tones be discriminated from another? Can enough easily discriminable patterns be devised to provide for all the alternative kinds of information in one signaling system? It is quite possible than an auditory pattern composed of alternating frequencies would provide a very positive kind of signaling system.

Stimulus Interactions

In a complex auditory signaling system, it will be necessary to change more than one physical parameter, and the observer will have to make judgments about more than one physical parameter. But judgments are made about psychological parameters, and frequently changing one physical variable changes more than one psychological variable. It is necessary, then, to know about such effects so they can be avoided when necessary.

It has long been known that the loudness of a tone depends on its frequency (**4**) and that the pitch of a tone depends on its intensity (**29**).

We know that the loudness of a tone changes with duration. Fig. 4 illustrates

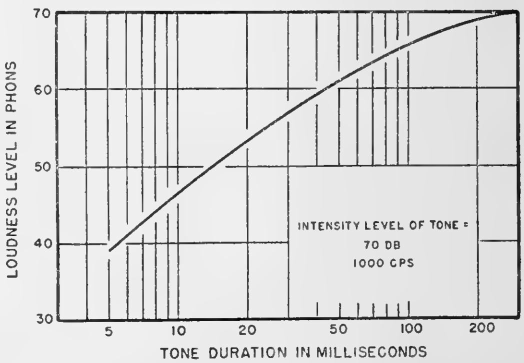

Fig. 4. The effect of tone duration on loudness
The ordinate values are the equivalent loudness levels of tones of various durations. The intensity and frequency of the steady tone are indicated. (After Munson, **25**)

how great this change may be, and is adapted from a report by Munson (**25**). Notice that the equivalent intensity of a tone may vary by as much as 30 db when the duration of the tone is changed from 5 to 200 milliseconds. Hughes (**20**), Garner and Miller (**15**), and Garner (**11**) have shown that the threshold of a tone varies inversely with its duration over a wide range.

Garner (**12**) has recently shown that the loudness of a series of tones increases with an increase in repetition rate, as illustrated in Fig. 5. In an earlier paper (**10**), he had also demonstrated that the threshold of a tone changes with a change in repetition rate.

An interesting thing about threshold changes was that the threshold was lower for faster repetition rates even at rates so slow that no real integration could be taking place. At faster rates, of course, it is reasonable to expect that there would be some increase in effective energy as more and more tones are added in a brief period of time.

Auditory Localization

One area of auditory research that has received considerable attention throughout the years has been the problem of auditory localization. Still, much of the large literature has proved almost useless for predicting

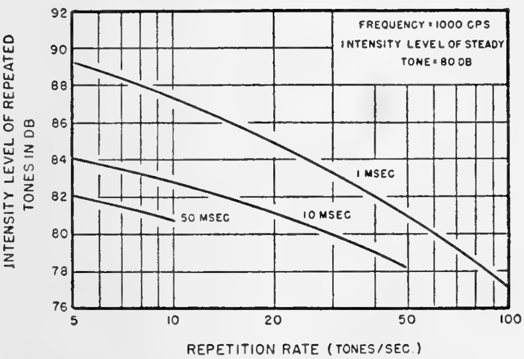

Fig. 5. The effect of repetition rate of short tones on loudness

The ordinate values indicate the intensity of the repeated tones necessary for them to have a loudness equal to that of a steady tone with an intensity level of 80 db. These curves are equal loudness contours. The duration of the repeated tones is shown for each curve. (After Garner, **12**)

how well any auditory signaling system will work. The main reason for this is that most of the research has been concerned primarily with understanding how we localize sound sources in real life, rather than with the problem of simulating an apparently located sound source.

Localization of Sound Sources

Stevens and Newman (**32**) have provided data on our ability to locate a source of a pure tone in open air. The error of the localization is shown in Fig. 6. Localization is poorest in the area of 2000 to 4000 cps, but is much better at the lower and higher frequencies. These authors postulate that this effect is due to the fact that interaural intensity cues are best at the higher frequencies, where the shading effects of the head are greatest; and that interaural phase cues are best at low frequencies, where the wave length of a tone is long enough for phase differences to have some meaning.

The literature has shown without question that both intensity differences and phase differences between the ears can serve as cues for the localization of a sound source. In addition, of course, differences in time of

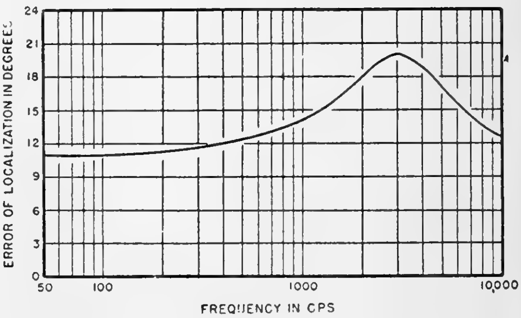

Fig. 6. Localization errors for pure tones in open space

The ordinate shows the average error made in determining the direction of a pure tone source at different frequencies. (After Stevens and Newman, **32**)

arrival of tones can serve as cues for non-continuous tones.

Simulated Localization

When the problem of auditory signaling is mentioned, usually the first thing that comes to mind is the use of some indication of lateral displacement of a sound source. If time, intensity, and phase differences between the two ears are the cues we use in localizing a sound source, then it should be simple to produce an apparent displacement of a sound source by stimulating the two ears differently in one of these respects. Unfortunately, it is not as simple as that. The simulated localization is never as effective as real localization, at least when only one type of stimulus cue is used.

Intensity Differences. Both Garner (**9**) and Ford (**8**) have shown that the normal person's ability to equate the intensities of tones in the two ears is not nearly as good as would be predicted from the data on localization. A much greater intensity disparity between the two ears is needed in order to produce an apparent displacement than seems reasonable. In addition, the experiments by Garner showed that many observers have two ears that do not agree about the loudness of a given tone intensity. For example, an intensity of 60 db in one ear may be equally loud to an intensity of 70 db in the other ear. The ears are, in a sense, off balance in loudness when both ears are stimulated with the same intensity. This same effect had also been shown incidentally by Forbes *et al.* (**7**).

Phase Differences. Phase differences at low frequencies have shown a somewhat better picture in some respects. The most recent study of threshold of detection for phase differences was reported by Hughes (**19**). For experienced subjects, the threshold for the detection of a shift of the tone from the median plane was about 20 degrees phase difference at the lower frequencies. This figure was closer to 40 degrees for inexperienced observers, however. Nevertheless, these figures agree with the Stevens and Newman data much better than the data for intensity differences do. But even Hughes' data indicate how unreliable is such a simulated localization. For example, his experienced subjects detected an abrupt change from zero degrees phase difference to 90 degrees phase difference only 80 percent of the time, at the lower frequencies. There seems little question that attempts to simulate apparent localization with the use of simple phase or intensity differences between the ears will not produce the positive stimulus required.

Combination Cues. The final answer to the localization problem, on which more research is definitely needed, may be that good simulated localization can be produced only by the combination of two or more of the cues. In an extensive series of experiments, Langmuir *et al.* (**22**) simulated a sound source by practically duplicating the interaural differences which occur in real situations. The task of their observers was to make the sound appear to be in the median plane, which they did by adjusting the apparatus. Thus they heard the sound appear to change location, and eventually "centered" it. Their results show an ability to localize the sound source in the median plane with a precision greater than that shown in the Stevens and Newman experiments.

Dynamic Cues. This greater precision of localization suggests that dynamic, or moving, cues are much better than static cues. In the Stevens and Newman experiment, the observers made a static localization; i.e., they determined the azimuth location of a tone source which was stationary. In the Langmuir *et al.* experiment, the subjects were able to listen to all simulated positions, and hear the changes as they manipulated the apparatus. Thus they had many of the dynamic cues which are available in real situations—cues which depend on movement of the sound source or movement of the head and body with respect to the source. Wallach (**36**), too, has emphasized the dynamic nature of sound localization in real situations.

Whatever the final outcome of this type of research, it is clear that more research is needed before it will be possible to simulate good localization of sounds. It may be that the best simulation, or the best artificial localization, will occur with unrealistic devices. Langmuir and his associates, for example, found that localization was very good when the observers could adjust the sounds to give a binaural null at a position corresponding to straight ahead. This method does not provide a realistic apparent sound source, but does provide accurate positioning on the part of the operator.

Stimulus Interactions: Phase Effects. The sound localization experiments will in themselves make more research necessary on the problem of stimulus interactions. Hirsh (**17**), for example, has recently shown that the masked threshold of a tone depends on the interaural phase relations of the binaural tone. Fig. 7 illustrates the amount of this effect for one condition. The use of interaural phase differences to simulate localization, then, would have to take this changed threshold into account. Experiments are needed to determine whether the same kind of a change can occur for the loudness of a tone as well. It is quite likely that both the pitch and the loudness of a tone may also depend on the interaural phase relations.

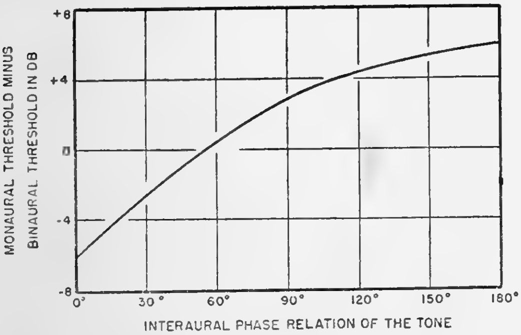

Fig. 7. The effect of interaural phase relations on the binaural masked threshold

The ordinate shows the threshold of a 200 cps binaural tone relative to the monaural threshold, at different interaural phase relations. When the ordinate value is negative, the binaural threshold is higher than the monaural threshold, indicating inhibition. Positive values indicate summation. (After Hirsh, **17**)

Another type of interaction has interesting possibilities. Hughes (**18**) demonstrated that interaural loudness matching is more precise when the two ears are stimulated 180 degrees out of phase than when the two ears are stimulated with the same phase relations. Unfortunately, his data are somewhat limited. More research along these lines, however, might prove very fruitful.

Other Auditory Illusions

Simulated sound localizations are auditory illusions. The listener hears something which does not really exist, because the stimuli with which he is presented are misinterpreted. Auditory localizations are not the only possible auditory illusions which can be produced, however. Illusions of movement can be produced, as demonstrated in the Flybar experiments. A considerable amount of research is needed, however, on illusions of movement and on any other possible type of illusion. There are practically no leads for research in this area, and yet data here could have great bearing on the future of auditory signals.

Psychological Scaling

If auditory signals are to be used to transmit quantitative information, i.e., information about real quantities, then the auditory signal must itself have some quantitative attribute. For example, it will not be enough to know that one tone is louder than another; it must also be possible to know *how much* louder. A psychological scale for loudness has been established by Stevens (**30**), and one for pitch by Stevens, Volkmann, and Newman (**33**). These data on pitch and loudness scales are adequate for many purposes, but do not provide answers for all the questions. For instance, is the pitch scale the same for tones having a high harmonic content as for pure tones? Are the scales affected by the introduction of a constant reference tone? The following example illustrates one kind of problem that needs to be solved before quantitative information can be transmitted with auditory signals.

Suppose that a reference tone is followed by a comparison tone, and the intensity of comparison tone with respect to that of the reference tone indicates some quantitative information. Zero quantity would be represented by equality of the two tones. A quantity of five, say, would be represented by plus or minus three db. How much would 10 then be? Is the quantity of plus 10 best represented as twice the loudness of the tone that indicates a quantity of plus five, or by twice the difference in loudness between the standard and comparison tones? In loudness units, let us say that the reference tone has a loudness of 1000 millisones.

When the comparison tone has a loudness of 2000 millisones we will arbitrarily give it a value of plus five. Now, should plus ten be represented by 4000 or 3000 millisones? In the former case, the plus 10 value is twice that of the plus five value; in the latter case the difference between standard and comparison is twice as great. In other words, even though we assume that the present loudness scale can predict what these quantities are for a particular intensity, we still do not know which kind of scale is more meaningful in a communications situation.

Auditory Counting

One other area of research which could provide useful information for signalling systems is that concerned with the counting of a series of tones or impulses. How fast can a series of tones be repeated if the listener is expected to get an accurate count of the total number? Can number patterns be learned, even though accurate counts are not possible? With a limited kind of information, and a limited quantitative scale, very accurate information could be transmitted automatically with series of impulses or tones. We need to know how well this type of signal can be used, but there is no information available at present.

The Psychomotor Problem

Throughout this discussion we have talked almost entirely about the purely perceptual problems. We have been concerned with what can be heard. In the main, these are the really important and basic problems, but they are not the only problems.

Motor Coordination

The problem of auditory signalling is basically a psychomotor problem, at least in most situations. The auditory signal is provided because somebody is supposed to take some action. In both radio range and Flybar operation, for example, a continuous motor adjustment is required as a result of the information received from the auditory signal.

The problems of audio-motor coordination of this kind have been investigated very little. A program of research analogous to the work done on visual-motor performance is necessary. There is, for example, a considerable body of research on eye-hand coordination, but practically none on ear-hand coordination.

Audio-Visual Comparisons

One of the best reasons for wanting auditory signals is that the eyes are asked to do more and more work in many of our highly complex modern equipments. Some intersensory comparisons are necessary to determine just what type of situation can be handled best with auditory signals. In many cases such research would almost become a study of the breakdown of the visual perception process, but they would be useful to tell us when auditory signals could be most valuable.

As an illustration of this kind of problem, one typical experiment might be concerned with measuring the precision of an eye-hand coordination involving two movements in response to information provided from two dials or meters. As these dials are moved farther and farther apart what happens to the coordination? At what distance can the operator no longer follow both meters? Similar comparisons would be made with one and two auditory signals, either alone or in conjunction with one visual signal.

Summary

Auditory signals are a form of display, but are usually characterized by being discrete in time, and some action is assumed on receipt of the information. Whenever a communications situation is concerned with a limited kind of information, an auditory signal can be used to provide a fast indicator of what to do. They are likewise useful as a substitute for visual displays when the visual load on an operator becomes too great. One of the chief advantages of the auditory signal is that the transmission of information can be completely automatic.

The radio range is a relatively simple form of auditory signal, and research on the use of the radio range involves simple intensity discrimination of tones. Flybar, on the other hand, is a complex form of auditory signal which provides three simultaneous signals for use by aircraft pilots in flight. These two systems of auditory signals have been used to illustrate the type of research which is done on these problems.

The major research requirement for auditory signals is basic research in the psychology of hearing. Of the many topics which are covered in that area, however, research on simple and complex discrimination, stimulus interactions, auditory localization, psychological scaling, and auditory illusions appears to be the most profitable from the point of view of possible application to the problems of auditory signalling.

References

1. DeFlorez, L. True blind flight. *J. aeronaut. Sci.*, 1936, **3**, 168–170.
2. Doughty, J. M. & Garner, W. R. Pitch characteristics of short tones. I. Two kinds of pitch threshold. *J. exp. Psychol.*, 1947, **37**, 351–365.
3. Doughty, J. M. & Garner, W. R. Pitch characteristics of short tones. II. Pitch as a function of tonal duration. *J. exp. Psychol.*, 1948, **38**, 478–494.
4. Fletcher, H. & Munson. W. A. Loudness, its definition, measurement and calculation. *J. acoust. Soc. Amer.*, 1933, **5**, 82–108.
5. Flynn, J. P., Goffard, S. J., Truscott, I. P., & Forbes, T. W. *Auditory factors in the discrimination of radio range signals.* Report from the Psycho-Acoustic Laboratory, Harvard Univ., 31 December 1945. OSRD Report No. 6292. (Available through Publication Board, U. S. Department of Commerce, Washington, D. C.)
6. Forbes, T. W. Auditory signals for instrument flying. *J. aeronaut. Sci.*, 1946, **13**, 255–258.
7. Forbes, T. W., Garner, W. R., & Howard, J. G. *Flying by auditory reference ("Flybar").* Report from the Psycho-Acoustic Laboratory, Harvard Univ., 1 June 1945. OSRD Report No. 5123. (Available through Publication Board, U. S. Department of Commerce, Washington, D. C.)
8. Ford, A. Dynamic auditory localization. 1. The binaural intensity disparity limen. *J. acoust. Soc. Amer.*, 1942, **13**, 367–372.
9. Garner, W. R. Accuracy of binaural loudness matching with repeated short tones. *J. exp. Psychol.*, 1947, **37**, 337-350.
10. Garner, W. R. Auditory thresholds of short tones as a function of repetition rates. *J. acoust. Soc. Amer.*, 1947, **19**, 600–608.
11. Garner, W. R. The effect of frequency spectrum on temporal integration of energy in the ear. *J. acoust. Soc. Amer.*, 1947, **19**, 808–815.
12. Garner, W. R. The loudness of repeated short tones. *J. acoust. Soc. Amer.*, 1948, **20**, 513–527.
13. Garner, W. R. Loudness and monaural loudness matching of short tones. (In preparation)
14. Garner, W. R. & Miller, G. A. Differential sensitivity to intensity as a function of the duration of the comparison tones. *J. exp. Psychol.*, 1944, **34**, 450–463.
15. Garner, W. R. & Miller, G. A. The masked threshold of pure tones as a function of duration. *J. exp. Psychol.*, 1947, **37**, 293–303.
16. Harris, J. D. Studies on pitch discrimination in masking. II. The effect of signal/noise differential. *J. acoust. Soc. Amer.*, 1947, **19**, 816–819.
17. Hirsh, I. J. The influence of interaural phase on interaural summation and inhibition. *J. acoust. Soc. Amer.*, 1948, **20**, 536–544.
18. Hughes, J. W. Binaural localization with two notes differing in phase by 180°. *Brit. J. Psychol.*, 1939, **30**, 52–56.
19. Hughes, J. W. The upper frequency limit for the binaural localization of a pure tone by phase difference. *Proc. roy. Soc.*, 1940, **B128**, 293–305.
20. Hughes, J. W. The threshold of audition for short periods of stimulation. *Proc. roy. Soc.*, 1946, **B133**, 486–490.
21. Karlin, J. E. *Auditory tests for the ability to discriminate the pitch and loudness of noises.* Report from the Psycho-Acoustic Laboratory, Harvard Univ., 1 August 1945. OSRD Report No. 5294. (Available through Publication Board, U. S. Department of Commerce, Washington, D. C.)
22. Langmuir, I., Schaefer, V. J., Ferguson, C. V., & Hennelly, E. F. *A study of binaural perception of the direction of a sound source.* General Electric Research Laboratory Report. (Available through Publica-

cation Board, U. S. Department of Commerce, Washington, D. C.)

23. Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. acoust. Soc. Amer.*, 1947, **19**, 609–619.

24. Montgomery, H. C. Influence of experimental technique on the measurement of differential intensity sensitivity of the ear. *J. acoust. Soc. Amer.*, 1935, **7**, 39–43.

25. Munson, W. A. The growth of auditory sensation. *J. acoust. Soc. Amer.*, 1947, **19**, 584–591.

26. Pepinsky, A. *Pitch as a function of the duration and extent of frequency modulation.* Ph.D. Thesis, State Univ. of Iowa, 1939.

27. Riesz, R. R. Differential intensity sensitivity of the ear for pure tones. *Phys. Rev.*, 2nd Series, 1928, **31**, 867–875.

28. Shower, E. G. & Biddulph, R. Differential pitch sensitivity of the ear. *J. acoust. Soc. Amer.*, 1931, **3**, 275–287.

29. Stevens, S. S. The relation of pitch to intensity. *J. acoust. Soc. Amer.*, 1935, **6**, 150–154.

30. Stevens, S. S. A scale for the measurement of a psychological magnitude: loudness. *Psychol. Rev.*, 1936, **43**, 405–416.

31. Stevens, S. S. & Davis, H. *Hearing.* New York: John Wiley, 1938.

32. Stevens, S. S. & Newman, E. B. The localization of actual sources of sound. *Amer. J. Psychol.*, 1936, **48**, 297–306.

33. Stevens, S. S., Volkmann, J., & Newman, E. B. A scale for the measurement of the psychological magnitude pitch. *J. acoust. Soc. Amer.*, 1937, **8**, 185–190.

34. Taylor, D. W. Learning telegraphic code. *Psychol. Bull.*, 1943, **40**, 461–487.

35. Turnbull, W. W. Pitch discrimination as a function of tonal duration. *J. exp. Psychol.*, 1944, **34**, 302–316.

36. Wallach, H. On sound localization. *J. acoust. Soc. Amer.*, 1938, **10**, 270–274.

Chapter 9

AUDITORY DISCRIMINATION IN SONAR OPERATION

W. D. NEFF AND W. R. THURLOW

University of Chicago and University of Virginia

Introduction

The surface craft or submarine obtains information about its environment through numerous channels: direct vision, radar, radio, and sonar. Of these sonar is of particular importance to the completely submerged submarine, to the surface craft in detecting submarines, and to either submarine or surface craft when it wishes to detect objects, such as mines, which are hidden under the water. Sonar makes use of sound waves which are transmitted readily through the water; it picks up the sound signals from the water, amplifies them, and analyzes them in various ways so as to present to the operator of sonar equipment useful items of information about objects and their movements in the sea around him.

Sonar systems may be divided into two basic types. *Listening systems* make use of the sounds emitted by objects in the water, e.g., the cavitation noise produced by the screws of a ship. *Echo-ranging systems* transmit pulses of sound into the water and make use of the reflected sound energy, or echoes, received from objects in the path of the transmitted pulses. In general, listening systems use sound frequencies in the sonic range, 20,000 cycles per second or less, the range in which transmission of sound through water is most efficient. However, some listening systems operate in the super-sonic range, above 20,000 cps. Echo-ranging systems, in order to control the direction of transmission of sound pulses, usually use frequencies in the super-sonic range.

With both listening and echo-ranging systems, the sound energy picked up by a hydrophone (underwater microphone) is changed into electrical energy, which is amplified and treated in one way or another—e.g. passed through selective filters—, then is changed back to sound energy, through a speaker or headphones, or to light energy, through a cathode-ray oscilloscope, and finally is presented to the sound operator.

In sonic listening systems, the operator receives most of his information aurally and the sounds he hears have a "real" quality, i.e., the turning of a ship's screws in the water sounds like the churning of water. In super-sonic listening systems, the sounds as presented to the operator after heterodyning have a more "artificial" character, are more like bursts of static or pulses of noise lacking the subtle quality differences of the sonic system.

The sounds heard in the typical echo-ranging system are primarily tonal but with a noise background. Echoes from the transmitted pulse are heterodyned so that the operator hears them as a somewhat irregularly modulated tone with a frequency of 800 cps (approx.). As a pulse of sound is transmitted, reverberations (echoes which come principally from the water surface and any small particles in the water) are first heard. The reverberations gradually decrease in intensity and echoes from objects at a distance are heard as more or less abrupt pulses of sound of slightly greater intensity and, in the case of moving objects, of different frequency from the background reverberations. The echoes and reverberations, both of which have a tonal character, are superimposed on a background of system noise and water noise, the latter produced primarily by movement of the hydrophone through the water.

In echo-ranging systems, much of the information received is presented to the sonar operator visually as well as aurally. In the present chapter, however, we shall be concerned with auditory discrimination and will mention only casually problems related to combined visual-auditory presentation. Research on auditory-visual displays is discussed in detail elsewhere in this book.

Auditory Tasks of the Sonar Operator

In the preceding introductory paragraphs, a very brief description of sonar systems and of the sounds which they deliver to the sonar operator has been given. Before discussing methods and results of auditory research in relation to sonar, the principal kinds of auditory discriminations required of the sonar operator will be outlined.

Listening Systems

The operator of the listening type of sonar gear must first of all detect a sound signal which appears in a noise background. The signal will usually be atonal in character and will differ from the background in one or more of these characteristics: loudness, rhythm, and quality. Occasionally, as in the case of the whining noise produced by machinery or by friction of the driving shaft of a ship's screws, a signal may have a tonal character.

Having detected a signal, the sonar operator must attempt to identify it and, in the case of a moving target, he must try to deduce its maneuvers. Identification will be made in terms of rate and quality of rhythmic patterns heard. For example, a large ship with relatively slow turning screws may produce a pulsating sound with a slow, heavily accented rhythmic beat, while a small, rapidly moving craft will produce a very rapid pulsing sound with less distinctly accented beats. Shaft whines or squeaks may aid in identification. Marine animals, such as croakers and crackling shrimp, produce distinctive sounds which lack the rhythm or periodicity of the sounds produced by a ship's screws.

Moving targets can be discriminated from non-moving targets by noting changes in relative bearing. Target maneuvers may be detected not only by plotting bearings of the target but also by changes in rhythmic patterns of sound produced by the target's screws. Change in target speed will be heard as increase or decrease in rate of rhythmic beat. Approach or withdrawal of the target will be heard as increase or decrease in intensity of the target-produced sounds. However, as a target with bow towards the listening ship turns away, presenting its stern, the sounds from the screws will usually become louder for a short time until the distance has increased somewhat. A change in quality may also become apparent. These changes occur because the sound waves produced by the screws are partially blocked by the hull of the bow-on vessel and will come with greater intensity and less loss of higher frequencies when the stern is towards the listening ship.

To summarize: the operator of sonar listening gear needs to make the following auditory discriminations:

1. Detection of a noise signal masked by noise background, the signal differing from the background in one or more of these characteristics: loudness, rhythm, quality.
2. Recognition of changes in a noise signal masked by a noise background, the usual changes being those of loudness, quality, and rate of rhythmic pulsation.

Echo-ranging Systems

Typically, the first task of the operator of echo-ranging sonar is to detect a short tonal pulse or "pip" which appears against a background composed of noise and, more predominantly, of a tone, irregularly modulated, and gradually decreasing in amplitude. The "pip" which is the echo from a target will differ from the background in loudness and, if from a target moving towards or away from the echo-ranging ship, in pitch. The change in pitch is due to the Doppler effect. Moving objects in the water may be distinguished from stationary objects both on the basis of relative bearing changes and Doppler.

If the target contacted is a moving ship, its maneuvers may be plotted from changes in relative bearing and changes in range. More important for our present discussion, however, are the maneuvers which can be detected directly by the sonar operator from the sounds which he hears. Due to the Doppler effect, if the target is headed toward the echo-ranging ship, the echoes from it will be higher in pitch than the reverberations; if it is headed away from the echo-ranging ship, the echoes will be lower in pitch; if it is beam-on (i.e., neither headed towards nor away from the echo-ranging ship), the echoes will be of the same pitch as the reverberations. Changes in the target ship's course will, therefore, be reflected in changes in the pitch differences between reverberations and echoes. For example, if the sonar operator is receiving echoes whose pitch is approximately the same as that of the reverberations and then the echoes gradually become higher in pitch, he knows that a beam-on target has turned towards him and is now bow-on.

Target aspect, i.e., the heading of the target with respect to the beam of sound striking it, and changes in target aspect are also reflected in terms of loudness, duration, and quality of the echoes heard by the sonar operator. A target at right angles to the sound beam striking it will return an echo which is louder, "sharper" in quality, and briefer in duration than the echoes from a target which is headed towards or away from the projector sending out the sound pulses. Furthermore, the echo from a target, whose stern is towards the echo-ranging ship, will differ from one which is bow-on because of the variations in the echo produced by the wake. Echoes from stern-on targets will usually be longer in duration and "mushier" in quality.

These examples will serve to illustrate the principal kinds of auditory discriminations required of the operator of echo-ranging sonar. They may be summarized as:

1. Detection of a tonal signal masked by a background composed of an irregularly modulated tone and of noise, the signal differing from the background principally with regards to loudness or pitch. The rate of change of loudness—the abruptness of the echo—is also a factor to be considered.

2. Recognition of changes in the tonal signals masked by the tonal and noise background, the changes being those of pitch, loudness, duration, and quality.

Methods of Investigation[1]

In basic auditory research three classes of methods are commonly used, namely, psycho-physical, physio-physical, and psycho-physiological. Theories of hearing have been based upon the results obtained by all of these methods. Immediate answers to the auditory problems encountered in sonar will undoubtedly come from psycho-physical studies. From a long-range point of view, however, it is desirable to extend our knowledge of the whole process of hearing—to know what physiological events are set off by changes in the physical stimulus and to know what physiological processes underlie the discriminatory responses made by the living organism.

Psycho-physical Methods

Psycho-physical methods are those in which the independent variables which are manipulated by the experimenter are various attributes of the physical stimulus, e.g., frequency, intensity, and complexity.

[1] This and the following section on Results of Relevant Research, etc. are reproduced, with minor changes, from a report prepared by one of the authors (W. D. N.) for the Panel on Underwater Acoustics, Committee on Undersea Warfare, NRC.

The dependent variables which the experimenter observes and measures are discriminatory responses of experimental subjects. Where human subjects are used these responses will as a rule be verbal reports; where lower animals are used, discriminations will be indicated by overt motor reactions. Standard procedures used in psycho-physical studies are in common use in the psychological laboratory. See, for example, Guilford (**21**), Stevens and Davis (**52**), and Thurstone (**56**).

Physio-physical Methods

In studying the physiological changes evoked by changes in the physical stimulus, the independent variables are again attributes of the physical stimulus, such as intensity, frequency, and complexity; but the dependent variables, in this case, are directly recorded physiological changes in the auditory system. The physiological changes which have been most commonly used in the past are the microphonic response of the cochlea and the action potentials of the auditory nerve or of the higher neural pathways (**5, 16, 52, 62**). Recently Békésy (**4**) has developed a technique which makes possible visualization of movement of cochlear structures. This method appears to have much promise. Experimental subjects in physio-physical investigations will usually be lower animals.

Psycho-physiological Methods

Finally, it is necessary to relate physiological changes and psychological phenomena. The independent variables will be experimentally produced alterations of the physiological mechanism of hearing, i.e., of the middle ear, inner ear, or neural structures; and the dependent variables will be discriminatory responses of the experimental subjects.

As in physio-physical studies, animals will necessarily be used as experimental subjects in most psycho-physiological experiments. Methods have been developed which make possible accurate determination of the ability of animals to discriminate changes in auditory stimuli (**6, 13, 29, 45**). Some information may also be derived from clinical studies of man in which variation of the auditory structures is produced by disease or other natural causes rather than by the experimenter (**12, 20**).

Results of Relevant Research and Suggestions for Future Investigations

Experiments Using Noise Stimuli

Auditory theory has been founded, in large part, upon experiments in which pure-tone stimuli were used. Field investigations, such as those in underwater acoustics, indicate the inadequacy of present theory in dealing with problems relating to complex sounds. An extension of basic research to include a detailed analysis of how the auditory system functions when activated by stimuli of complex frequency spectra is, therefore, suggested.

In laboratories which have done research on auditory problems related to sonar, noise stimuli used in experiments more often than not have been designed to simulate as exactly as possible the actual noises which occur during operations at sea. For example, careful film or disc reproductions have been made of the cavitation noise produced by a ship's screws and of the ambient water noise. For basic studies, a constant and controllable noise source is required. Therefore, thermal (or "white") noise in which all frequencies are present and the spectrum of sound is continuous is to be preferred.

Determination of Psycho-physical Relationships

Using white noise stimuli, psycho-physical experiments which have been performed with pure tones can be paralleled, and changes in the psychological phenomena of loudness and pitch as a function of such variables as band width, intensity, and region of the frequency spectrum can be determined. Measures of the differential thresholds of loudness and pitch as functions of these same variables may also be obtained.

A number of psycho-physical studies using white noise stimuli have been done by the Psycho-Acoustic Laboratory, Harvard University. For example, recorded tests which measure ability to discriminate the pitch and loudness of noises have been developed by Karlin (**30**). Preliminary results on the use of the loudness discrimination tests in the selection of sonar operators have been reported by Harris (**23**). Miller (**38**) has measured sensitivity to changes in intensity over a wide range of intensities. He has found that:

> The just detectable increment in the intensity of the noise is of the same order of magnitude as the just detectable increment in the intensity of pure tones. For intensities more than 30 db. above the threshold of hearing for noise the size in decibels of the increment which can be heard 50 percent of the time is approximately constant (0.41 db.).

He also found that, as for pure tones, just noticeable differences at different intensity levels are not equal in subjective magnitude, the j.n.d. at low intensity levels appearing smaller than at high levels. (See below for discussion of the relation of these results to the phenomenon of masking.)

Measurement of Physiological Events

Physio-physical experiments which may be done using white noise stimuli include determination of changes in cochlear microphonics and nerve potentials as a function of the different variables of the physical stimulus.

Effect of Physiological Changes on Hearing of Noise Stimuli

Psycho-physiological experiments which may be performed will be concerned with changes in hearing of noise stimuli as a result of alteration of the physiological mechanism. For example, absolute thresholds of hearing for pure tones and for different bands of white noise can be measured before and after lesions of the cochlea or auditory nervous system.

To date, there are no reports available of research projects that might be classified under the above two headings.

Masking Experiments

Since one of the most important tasks of the sonar operator is the detection of an auditory signal which is more or less masked by a background of sound, studies of masking are of major practical importance in underwater acoustics. Moreover, it is in this area, the study of masking phenomena, that the field research appears most suggestive to the experimenter in the university laboratory.

Beginning with the classical study of Wegel and Lane (**60**) we have had numerous investigations of the masking of one pure tone by another. With the exception of the experiments on masking of speech by noise, which were carried on for the most part in the Bell Telephone Laboratories, there were almost no studies prior to World War II of the masking effects of noise on pure tones or of noise background on noise signals. The work of the laboratories investigating problems of underwater acoustics and of other laboratories studying the problems of communication in noise have emphasized the lack of knowledge in this area.

Masking of Pure Tones by Noise

The Harvard Psycho-Acoustic Laboratory has recently published the results of a series of experiments on the masking of pure tones by white noise (**26**). Tests were made at eight noise levels, ranging from 20 to 90 decibels. Results were used to determine two basic functions:

> (a) The critical band width of a masking noise, i.e., the ratio, in decibels, between the level of a pure tone and the level per cycle of the noise that is just able to mask the tone.
> (b) The function relating the amount of

masking to the effective level of the masking noise.[2]

Related studies have been reported by Harris (**24**) and Flynn, Truscott, and Newman (**14**). Detectability as a function of signal duration has been investigated by the Sonar Data Division of the University of California Division of War Research, U. S. Navy Electronics Laboratory (**70**) in a series of experiments in which the signal was a short tonal pulse and the masking background was thermal noise.

Further psycho-physical investigations of this sort are indicated. In addition, physio-physical and psycho-physiological experiments are required. As an example of the kind of information that may be obtained from such experiments, an experiment by Galambos and Davis (**17**) may be cited. These investigators found that the spontaneous discharge, which may be recorded from a single nerve fibre of the auditory nerve,[3] can be stopped by certain tones and noises. Furthermore, the discharge elicited by certain tones can be reduced or abolished by the simultaneous presentation of another tone or noise. This neural "inhibition" may play an important role in masking and its discovery must lead us to question the adequacy of the commonly accepted theory that masking occurs because the masking tone or noise activates the region of the cochlea and the nerve fibers ordinarily activated by the masked tone. It appears that masking is a more complex phenomenon physiologically than has usually been assumed, and that extensive experimentation using the most refined recording techniques must be done in order that we may understand the interaction which apparently occurs peripherally in the auditory nervous system and, undoubtedly, in the higher auditory centers as well.

On the psycho-physiological level further data as to the analytical function of the auditory system can be obtained by experiments in which the masking effects of bands of noise on pure tones is determined in animals experimentally deprived of part of the sensory or neural elements of hearing. For example, will a band of frequencies, whose individual components are not heard because of experimental damage to the cochlear hair cells or auditory nerve, mask the hearing of frequencies unaffected by the lesion? Studies which have been done using pure tone stimuli (**32, 39, 63**) may be paralleled with the added variable of a masking noise background.

Masking of Noise Signals by Noise Background

Experiments similar to those which have been described or suggested under the heading of "masking of pure tones by noise" should also be done using noise signals instead of pure tones. The degree of masking as dependent upon such variables as band width and modulation may be investigated. Only a few studies of this nature have as yet been undertaken. Experiments on the masking of a noise signal, simulating the underwater sounds produced by a ship's screws, have been reported by the University of California Division of War Research (**68**) by Harris (**22**), and by Snow and Neff (**50**). Miller (**38**) has suggested that the results obtained in his experiments on sensitivity to changes in white noise may be regarded as measures of the masking of noise by noise. He points out that the "functions which describe intensity discrimination also describe the masking by white noise of pure tones and of speech," and that the determination of differential sensitivity to intensity may be considered as a special case of the more general masking experiment.

[2] Quoted from author's summary.

[3] In a recent note in *Science* (**18**), Galambos and Davis present evidence which indicates that the potentials recorded in these experiments were from the cell bodies of second-order neurons rather than from primary fibers in the auditory nerve. This new finding does not affect the discussion presented here.

Modulation

Problems encountered in studying the detection of echoes against a background of reverberations (**70**) suggest the need for experiments to determine the effects of amplitude, frequency, and phase modulation of pure tones upon the physiological processes and psychological phenomena of hearing.

The problem of modulation in relation to hearing has been discussed by Stevens and Davis (**52**) and a number of experiments cited (**42, 48, 54, 61, 64, 66**). These must be considered as preliminary only in view of the many variables yet to be explored. Investigations should be extended to the measurement of effects of modulation of complex sounds as well as of pure tones.

Hearing of Short Pulses of Sound

In certain types of modern echo-ranging gear the use of extremely short pulses of sound and the consequent short echoes returned by targets calls attention to a number of auditory discrimination problems which have not been intensively investigated.

Stevens and Davis (**52,** pp. 100–106 and pp. 154–159) have reviewed experiments (prior to 1938) in which short impulses of sound were used in studying auditory function. In their review are included psychophysical studies on pitch as a function of duration (**8, 53**), the effect of duration upon the difference limen (DL) for frequency discrimination (**3**), the threshold of successiveness—i.e., the minimum time of separation of two sounds in order that they appear as two rather than as one (**9, 55**) loudness as a function of duration (**2, 37, 51**), loudness as a function of the form of the pressure wave of the sound impulse (**10, 51**), and loudness as a function of the frequency of repeated impulses (**51**).

In a recent experiment Turnbull (**58**) has examined the effect of stimulus duration upon pitch discrimination for pure tone stimuli of 128, 1024, and 8192 cps. He has found that decreasing the duration of a tone to be compared with a standard tone of fixed duration has slight effect upon the relative difference limen $\left(\frac{\Delta F}{F}\right)$ until the length of the comparison tone has been reduced to 0.1 to 0.5 seconds, depending upon the frequency. At this point the difference limens begin to increase markedly. In the region of .01 to .03 seconds, accuracy reaches a virtual zero.

Turnbull also found that for a given duration of the comparison tone, decrease in intensity reduced the accuracy of pitch discrimination, and that this reduction was greater for tones of relatively short duration.

Garner and Miller (**19**) have investigated the differential sensitivity to intensity as a function of the duration of the comparison tone. For tones of 500 and 1000 cps, the former at sensation levels of both 40 and 70 db and the latter at 40 db only, they found that the relative difference limen $\left(\frac{\Delta I}{I}\right)$ was approximately constant for durations above 300 milliseconds. Below 300 milliseconds the reduction in accuracy of discrimination was much less rapid for the 500 cycle tone at a sensation level of 70 db than at 40 db. However, additional data should be obtained before any broad generalization about the effects of intensity level is made.

To further our knowledge of auditory function, numerous psychophysical studies in which duration of stimulus is a variable need to be performed. The results of Turnbull and of Garner and Miller should be extended to include tests with a wider range of frequencies and intensities. Functions paralleling those for tonal stimuli should be determined for noise stimuli. It would also be of interest to compare the loudness-duration functions for different intensity levels, intensity being measured as rms, peak and average pressures.

No attempt has been made to measure systematically the effects of short impulses, tonal or non-tonal, upon physiological mechanisms of the inner ear or auditory nervous system. A few observations have been made incidentally in the course of experiments aimed at the solution of other problems (**52**, pp. 327–332 and pp. 386–411). Some very suggestive results have more recently been reported by Galambos and Davis in their studies of the activity of single fibers of the auditory nerve (**16, 17**).[4] They have noted that "equilibration," the decrease in voltage output of the auditory nerve following onset of sound stimulation, occurs in the single fiber preparation and is essentially complete within a few tenths of a second (**16**). They have also made observations on the "inhibitory" or "masking" effect of a noise which consisted of a series of sharp clicks (**17**). Brief atonal and tonal impulses have also been used in investigating the response of the higher auditory centers, especially the auditory cortex (**1, 5, 35, 54, 57**). Further experiments with controlled variation of duration and wave form of impulses should contribute to our understanding of the function of the auditory nervous system. Such experiments should include not only the recording of electrical activity of the nervous pathways and centers but also the measurement by behavioral methods of discriminatory ability before and after ablation of parts of the system.

Auditory Fatigue

By auditory fatigue is meant temporary impairment of hearing sensitivity resulting from stimulation of the ear by sound.[5] Experiments which have reported fatigue effects have usually used intensities of sound some 100 db re 10^{-16} watts/sq. cm. (**31, 34, 40, 41, 43, 44, 46, 50, 67, 69**). The sonar operator, during his ordinary duties, is probably seldom subjected to sounds of this intensity, and auditory fatigue may not be a factor of great importance in determining his proficiency. However, in view of such experiments as those of Rawdon-Smith (**43, 44**), in which he found threshold shifts as high as 56 db resulting from stimulation by pure tones at 100–110 db above threshold and changes in differential sensitivity for frequency following stimulation by pure tones 70–100 db above threshold, it appears within the realm of possibility that repeated exposure of the ear to fairly loud sounds such as encountered by the sonar operator in echo-ranging may affect the sensitivity to minimal intensity or differential sensitivity to frequency and intensity. Further information on the 'fatigue' or 'inhibitory' effects of pure tones and noises, including intensities in the range 60–100 db, needs to be acquired. Brief impulses as well as steady sounds should be used and the effect of repeated exposures without opportunity for recovery from any losses which appear should be examined.[6]

Possible Techniques of Improving Auditory Discrimination

The ability of the sonar operator to detect the signal representing a target can be improved in two ways: (1) by advance in the

[4] See footnote 3.

[5] Rawdon-Smith (**44**) has suggested that the term 'experimental deafness' rather than fatigue be used to describe the temporary hearing losses which he found to occur in human subjects following brief exposures to tonal stimuli in the range 100–110 db. above threshold. He also suggests the term 'auditory inhibition' to refer to the part of this deafness which appears to be due to a central nervous system phenomenon. Wever (**62**), in a review of experiments on electrical response of the cochlea and auditory nervous system, suggests that "in view of the extreme intensities (of sound) required (to produce depression in the electrical responses), and the long period taken for recovery, it would seem more appropriate to regard the effect as one of injury rather than as fatigue in the usual sense," and that "it is probable that stimulation deafness is a later and irreversible stage of the same injury process."

[6] See Davis, Morgan, *et al.* (**67**). These investigators found no cumulative effects when ears of human subjects were repeatedly exposed to intense tones and noise 110–130 db re 10^{-16} watts/-cm^2. However, subjects were not re-exposed to the intense stimuli until recovery (as measured by the audiogram) from the previous exposure was complete.

design of equipment, and (2) by use of techniques which lower the discriminatory thresholds. The first of these has been a matter of major consideration on the part of the physicist and engineer. The second has received little direct attention, although a number of studies, aimed as a rule towards the solution of other problems, have produced results which deserve further consideration.

In a recent experiment, for example, Shaw, Newman and Hirsh (**49**) have found that:

> The absolute threshold for pure tones is lower when both ears are stimulated than when either ear is stimulated alone. The normal summation at threshold is a function of the relative sensitivity of the two ears. For a group of listeners with substantially normal hearing in both ears, the binaural threshold is from one to two decibels lower than the best monaural threshold.

Similarly, it has been found that smaller differences in frequency and in intensity can be detected when the stimuli are presented binaurally rather than monaurally (**52**, pp. 87 and 141–142). These results are in part supported by evidence from physiological experiments which indicates that the absolute threshold for intensity is raised by the destruction of one cochlea (**7**).

Several investigators (**15, 25, 11**) have reported improvement in auditory discrimination resulting from coincident visual stimulation (see **47** for a review of these experiments.) This phenomenon of intersensory facilitation is not well substantiated and is open to various interpretations. The possibility remains, nevertheless, that some real gains may be made by combined visual and auditory presentation of near liminal signals.

It has been a common observation on the part of researchers in audition that the ability to detect minimal intensities, to discriminate changes in frequency and intensity, and to recognize particular patterns of sound increases with practice. Studies such as that of Wyatt (**65**) on improvability of pitch discrimination are important additions to knowledge derived from more casual observations. Many additional investigations must be conducted before we shall know with any degree of certainty the limits of improvement, individual differences in improvability, the degree of transfer of training from one particular type of discrimination to another, the most effective training procedures, and the factors which account for improvement.

Significant improvement in auditory discrimination may sometimes be brought about by new and unexpected discoveries. Potentially promising is the recent observation that interaural phase relations affect the detectability of a tone heard against a background of noise (**27, 28, 33, 36**).

When a tone is presented to both ears against a background of noise, the tone is more easily heard if it is in phase at the two ears while the noise is out of phase (or vice versa) than when *both* the tone and the noise are either in phase or out of phase. For example, when a tone of 200 cps is led to the two ears along with white noise, the detectability of the tone is improved by as much as 13 db merely by reversing the interaural phase of the tone and leaving the noise alone.

It is even possible to improve the detectability of a tone simply by reversing the connections to one earphone, thereby making the interaural phase of both the tone and the noise 180°.

These laboratory findings may or may not prove applicable to sonar listening. The important point is that through the prosecution of vigorous research in university laboratories new and unsuspected facts and principles will continue to emerge in a steady stream. Some of these discoveries will have a revolutionary impact on the military art.

References

1. Ades, H. W., & Felder, R. The acoustic area of the monkey (Macaca mulatta). *J. Neurophysiol.*, 1942, **5**, 49–54.

2. Békésy, G. v. Zur Theorie des Hörens. Über die Bestimmung des einem reinen Tonempfinden entsprechenden Erregungsgebietes der Basilarmembran vermittels Ermüdungserscheinungen. *Physik. Z.*, 1929, **30**, 115–125.
3. Békésy, G. v. Zur Theorie des Hörens. Über die eben merkbare Amplituden- und Frequenzänderung eines Tones. Die Theorie der Schwebungen. *Physik. Z.*, 1929, **30**, 721–745.
4. Békésy, G. v. On the elasticity of the cochlear partition. *J. acoust. Soc. Amer.*, 1948, **20**, 227–240.
5. Bremer, F. & Dow, R. S. The cerebral acoustic area of the cat. A combined oscillographic and cytoarchitectonic study. *J. Neurophysiol.*, 1939, **2**, 308–318.
6. Brogden, W. J. & Culler, E. A. Device for the motor conditioning of small animals. *Science*, 1936, **83**, 269–270.
7. Brogden, W. J., Girden, E., Mettler, F. A., & Culler, E. Acoustic value of the several components of the auditory system in cats. *Amer. J. Physiol.*, 1936, **116**, 252–261.
8. Bürck, W., Kotowski, P., & Lichte, H. Der Aufbau des Tonhöhenbewusstseins. *Elek. Nachr.-Techn.*, 1935, **12**, 326–333.
9. Bürck, W., Kotowski, P., & Lichte, H. Die Hörbarkeit von Laufzeitdifferenzen. *Elek. Nachr.-Techn.*, 1935, **12**, 355–367.
10. Bürck, W., Kotowski, P. & Lichte, H. Die Lautstärke von Knacken, Geräuschen und Tönen. *Elek. Nachr.-Techn.*, 1935, **12**, 278–288.
11. Child, I. L., & Wendt, G. R. The temporal course of the influence of visual stimulation upon the auditory threshold. *J. exp. Psychol.*, 1938, **23**, 108–127.
12. Crowe, S. J., Guild, S. R., & Polvogt, L. M. Observations on the pathology of high-tone deafness. *Bull. Johns Hopkins Hosp.*, 1934, **54**, 315–379.
13. Culler, E., Finch, G., Girden, E. & Brogden, W. Measurement of acuity by the conditioned response technique. *J. gen. Psychol.*, 1935, **12**, 223–227.
14. Flynn, J. P., Truscott, I. P. & Newman, E. B. Intensity discrimination as a function of signal-to-noise ratio and intensity level. *Amer. Psychologist.*, 1946, **1**, 277. (Abstract)
15. Freund, L., & Hofmann, L. Licht und Hören. *Med. Klinik*, 1929, **25**, 226–228.
16. Galambos, R. & Davis, H. The response of single auditory-nerve fibers to acoustic stimulation. *J. Neurophysiol.*, 1943, **6**, 39–57.
17. Galambos, R., & Davis, H. Inhibition of activity in single nerve fibers by acoustic stimulation. *J. Neurophysiol.*, 1944, **7**, 287–303.
18. Galambos, R., & Davis, H. Action potentials from single auditory-nerve fibers? *Science*, 1948, **108**, 513.
19. Garner, W. R., & Miller, G. A. Differential sensitivity to intensity as a function of the duration of the comparison tone. *J. exp. Psychol.*, 1944, **34**, 450–463.
20. Guild, S. R. Tone localization in the cochlea II. Discussion from the point of view of studies on human temporal bones. *Ann. Otol., Rhinol., Laryngol.*, 1935, **44**, 738–753.
21. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
22. Harris, J. D. *Preliminary report on propeller noise discrimination.* Report of the Medical Research Department, U. S. Submarine Base, New London, Conn., December 26, 1944.
23. Harris, J. D. *Functions of loudness discrimination in submarine sonar operation. Interval report.* Bureau of Medicine and Surgery, Research Division, Project X-53. Issued by Medical Research Department, U. S. Submarine Base, New London, Conn. April 5, 1945.
24. Harris, J. D. Pitch discrimination in noise. *Amer. Psychologist*, 1946, **1**, 276–277 (Abstract).
25. Hartmann, G. W. The facilitating effect of strong general illumination upon the discrimination of pitch and intensity differences. *J. exp. Psychol.*, 1934, **17**, 813–822.
26. Hawkins, J. E. *The masking of pure tones and speech by white noise.* Part I, OSRD Report No. 5387. (Psycho-Acoustic Laboratory, Harvard University, October 1, 1945.)
27. Hirsh, I. J. Binaural summation and interaural inhibition as a function of the level of masking noise. *Amer. J. Psychol.*, 1948, **61**, 205–213.
28. Hirsh, I. J. The influence of interaural phase on interaural summation and inhibition. *J. acoust. Soc. Amer.*, 1948, **20**, 536–544.
29. Kappauf, W. E. The application of conditioning methods to the study of discrimination and the measurement of differential thresholds in animals. *J. Psychol.*, 1943, **15**, 129–135.
30. Karlin, J. E. *Auditory tests for the ability to discriminate the pitch and loudness of noises.* OSRD Report No. 5294 (Psycho-Acoustic Laboratory, Harvard University, August 1, 1945.)

31. Kobrak, H. G., Lindsay, J. R., & Perlman, H. B. Experimental observations on the question of auditory fatigue. *Laryngoscope*, St. Louis, 1941, **51**, 798–809.

32. Kryter K. D., and Ades, H. W. Studies on the function of the higher acoustic nervous centers in the cat. *Amer. J. Psychol.*, 1943, **56**, 501–536.

33. Langmuir, I., Schaefer, V. J., Ferguson, C. V., & Henelly, E. F. A study of the binaural perception of the direction of a sound source. OSRD Report No. 4079 (NDRC 6.1-1840), PB-31014, June, 1944.

34. Larsen, B. Investigations in the fatigue of hearing. *Acta Oto-laryng.*, Stockh., 1942, **30**, 525.

35. Licklider, J. C. R., & Kryter, K. D. Frequency-localisation in the auditory cortex of the monkey. *Fed. Proc. Amer. Soc. Exp. Biol.*, 1942, **1**, 51.

36. Licklider, J. C. R. The influence of interaural phase relations upon the masking of speech by white noise. *J. Acoust. Soc. Amer.*, 1948, **20**, 150.

37. Lifshitz, S. Apparent duration of sound perception and musical optimum reverberation. *J. acoust. Soc. Amer.*, 1936, **7**, 213–221.

38. Miller, G. A. Sensitivity to changes in intensity of white noise and its relation to masking and loudness. *J. acoust. Soc. Amer.*, 1947, **19**, 609–619.

39. Neff, W. D. The effects of partial section of the auditory nerve. *J. comp. physiol. Psychol.*, 1947, **40**, 203–215.

40. Parker, R. C. The nature of fatigue in the auditory system. *Proc. Phys. Soc.*, Lond., 1938, **50**, 108–118.

41. Perlman, H. B. Acoustic trauma in man. Clinical and experimental studies. *Arch. Otolaryngol.*, 1941, **34**, 429–452.

42. Ramsdell, D. A. *The psycho-physics of frequency modulation.* Unpublished thesis, Harvard University.

43. Rawdon-Smith, A. F. Auditory fatigue. *Brit. J. Psychol.*, 1934, **25**, 77–85.

44. Rawdon-Smith, A. F. Experimental deafness. Further data upon the phenomenon of so-called auditory fatigue. *Brit. J. Psychol.*, 1936, **26**, 233–244.

45. Rosenzweig, M. Discrimination of auditory intensities in the cat. *Amer. J. Psychol.*, 1946, **59**, 127–136.

46. Rüedi, L. & Furrer, W. *Das akustische Trauma.* Basel: S. Karger, 1947.

47. Ryan, T. A. Interrelations of the sensory systems in perception. *Psychol. Bull.*, 1940, **37**, 659–698.

48. Seashore, C. E. (Ed.) *The vibrato. Univ. Ia. Stud. Psychol. Mus.*, 1932, **1**.

49. Shaw, W. A., Newman, E. B., & Hirsh, I. J. The difference between monaural and binaural thresholds. *J. exp. Psychol.*, 1947, **37**, 229–242.

50. Snow, W. B., & Neff, W. D. *Effects of airplane noise on listening with headphones.* Report of Columbia University Division of War Research at the U. S. Navy Underwater Sound Laboratory, Fort Trumbull, New London, Conn., Jan. 25, 1943.

51. Steudel, U. Über Empfindung und Messung der Lautstärke. *Hochfrequenztechnik und Electroakustik*, 1933, **41**, 116–128.

52. Stevens, S. S. & Davis, H. *Hearing.* New York: John Wiley and Sons, 1938.

53. Stevens, S. S. & Ekdahl, A. G. The relation of pitch to the duration of a tone. *J. acoust. Soc. Amer.*, 1939, **10**, 255. (Abstract)

54. Stowell, E. Z. & Deming, A. F. Aural rectification. *J. acoust. Soc. Amer.*, 1934, **6**, 70–79.

55. Strecker, F. Die Bemerkbarkeit von Einschwingzeiten. *Telegr.-u. Fernspr.-Techn.*, 1935, **24**, 1–5.

56. Thurstone, L. L. Psychophysical methods. In: T. G. Andrews (Ed.). *Methods of psychology.* John Wiley and Sons, New York, 1948.

57. Tunturi, A. R. Further afferent connections to the acoustic cortex of the dog. *Amer. J. Physiol.*, 1945, **144**, 389–394.

58. Turnbull, W. W. Pitch discrimination as a function of tonal duration. *J. exp. Psychol.*, 1944, **34**, 302–316.

59. Walzl, E. M., & Woolsey, C. N. Effects of cochlear lesions on click responses in the auditory cortex of the cat. *Fed. Proc. Amer. Soc. exp. Biol.*, 1942, **1**, 88.

60. Wegel, R. L., & Lane, C. E. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.*, 1924, **23**, 266–285.

61. Weinberg, M., & Allen, F. On the critical frequency of pulsation tones. *Phil. Mag.*, 1924, **47**, 50–62.

62. Wever, E. G. The electrical responses of the ear. *Psychol. Bull.*, 1939, **36**, 143–187.

63. Wever, E. G., & Neff, W. D. A further study of the effects of partial section of the auditory nerve. *J. comp. physiol. Psychol.*, 1947, **40**, 217–226.

64. Wingfield, R. C. An experimental study of the apparent persistence of auditory sensations. *J. gen. Psychol.*, 1936, **14**, 136–157.

65. Wyatt, R. F. Improvability of pitch discrimination. *Psychol. Monogr.*, 1945, **58**, No. 2, 1–58.

66. YOUTZ, R. E. P., & STEVENS, S. S. On the pitch of frequency modulated tones. *Amer. J. Psychol.*, 1938, **51**, 521–526.

67. *Final report on temporary deafness following exposure to loud tones and noise.* Prepared by the Committee on Medical Research of the Office of Scientific Research and Development. Boston, Mass., Dept. of Physiology, Harvard Medical School, 1943. Reported under contract OEM cmr-194 (H. Davis, C. T. Morgan, J. E. Hawkins Jr., R. Galambos, and F. W. Smith).

68. *Masking experiments.* Report No. U258. University of California Division of War Research at the U. S. Navy Radio and Sound Laboratory, San Diego, California. Sept. 15, 1944.

69. O.S.R.D. Report No. 889. July 31, 1942.

70. *Studies of the recognition of submarine echoes.* File report No. M431. Sonar Data Division, University of California Division of War Research at the U. S. Navy Electronics Laboratory, San Diego, Calif. Sept. 30, 1946.

PART IV.

COMMUNICATION

Chapter 10

VOICE COMMUNICATIONS: PERSONNEL AND PHRASEOLOGY

M. H. ABRAMS

Cornell University

I. The Problem of Communicating in a Noisy Environment

The spoken word is still the basic medium for coordinating the operation of a submarine. The aim of submarine communications is to pass the word with the maximum speed and the minimum error. Its principal hazard is the presence of interfering noises made by other talkers, by various pieces of submarine equipment, and above all, by the diesel engines.

During the last war the conditions on board a submarine were particularly favorable for voice communication. The standard approach and attack was made under water, on battery power, and this mode of propulsion is relatively quiet. Even under these favorable circumstances, however, submarine commanders returning from war patrols testified that errors and confusion in communications cost them targets and jeopardized the safety of their ships. Recordings made from battle circuits dramatically confirmed this testimony. These records show that some messages had to be repeated several times; others were misunderstood without detection of the error; some never got through at all.[1] In an attempt to remedy these failures, NDRC personnel, in 1944, assisted ComSubsLant in improvising a program for selecting and training communications personnel, and for standardizing voice procedures. The results of this project will be discussed below.

It must be anticipated that from now on the difficulty of passing the word quickly and accurately will become even more acute. Development of the "snorkel" apparently will make it possible to carry out a submerged attack using the diesel engines for motive power. Submarine diesels make a great deal of noise, a noise comparable in intensity, in some compartments, to the din in airplanes and tanks (see Fig. 1). Communication in these vehicles proved to be very difficult in the last war, but the problem on board a submarine will be more critical still. In planes, for example, messages are spoken over communication circuits which have relatively faithful frequency characteristics, utilize an external source of power for amplification, and terminate in earphones which help seal off ambient noise and pass the word directly into the ear of the listener. In a submarine, on the other hand, interior communications take place either through loudspeakers or over sound-powered phones. Loudspeakers are usually distant from the ear of the listener, and cannot be sealed off from interference. They are therefore extremely vulnerable to ambient noise, and are also subject to feedback—a phenomenon equally annoying to the listener and hazardous to intelligibility. Sound-powered phones depend on the voice as their only source of energy, and pass only a very narrow band of frequencies. What the listener hears is therefore both weak and distorted, and readily drowned out by competing noises.

But even if high-fidelity communication equipment is substituted for these inferior types, experience with communications in airplanes serves warning that it will remain

[1] One of these recordings was processed for use in talker-training courses under the title "Telephone Talking over a Submarine Battle Circuit During Actual Operations" (**24**, p. 22).

very difficult to pass messages in the face of masking noises with the desired speed and accuracy. The chain of communication incorporates people and speech, as well as telephone equipment, and, like other chains,

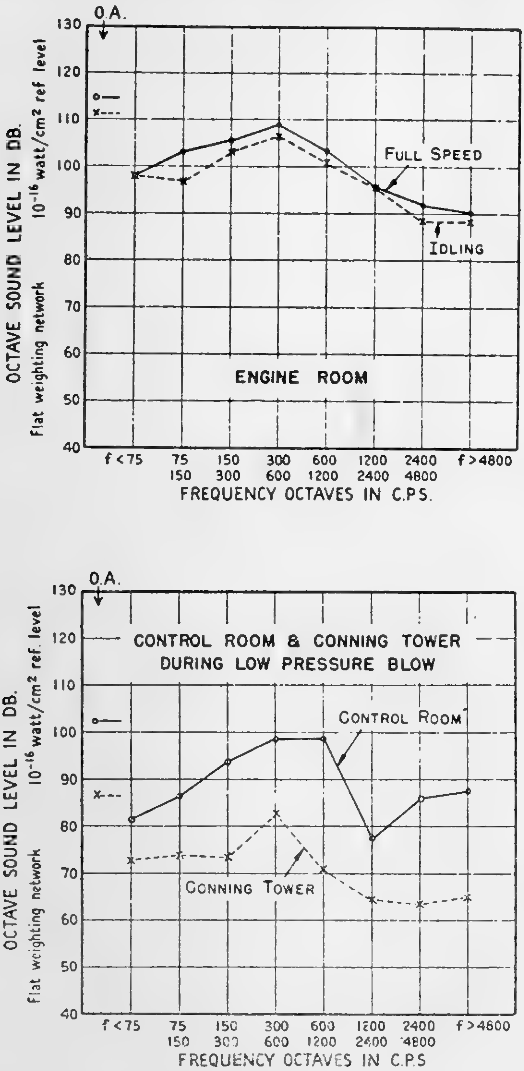

Fig. 1. Examples of noise spectra found in USS REDFISH. The symbol O.A. refers to overall sound pressure level. The sound pressure levels measured in successive bands one octave wide are shown by the points on the curves.

it is no stronger than its weakest link. Both the human and vocal links must be utilized to their utmost efficiency if we are to avoid the danger of critical failures in the operation of a submarine.

There are three ways in which the quality of the human and vocal elements in communication can be considerably improved:

1. Operators manning the more important communication posts can be selected for their inherent ability to speak intelligibly, as well as to interpret messages which have been altered by distortion and partially masked by noise.

2. All operators can be trained to speak more intelligibly, to use their equipment efficiently, and to improve their ability to interpret mutilated vocal signals.

3. All important messages can be standardized, so that—without departing too far from common usage—they are short, simple, easy to say, and easy to understand.

Considerable progress was made by the Submarine Service along all three of these lines in the late stages of the last war. This chapter will describe these developments, and, in addition, will summarize the existing knowledge about these matters, as an aid both in consolidating existing gains and in formulating a program for further research and improvement in the future.

It should be emphasized at the start that routine tests for acuity of hearing, standard criteria of "good speech," and conventional methods for rating and training public speakers, however valuable for the concerns of everyday life, are not valid for the peculiar conditions of military communication in noise. The criterion here is not "effectiveness," "pleasantness," nor any aesthetic quality, but the one crucial attribute: can what is said be heard and understood? For the most part knowledge of the attributes of speech, speakers, and listeners, conducive to efficient communication in noise, had to be discovered by special experimentation during the last war.

II. The Selection of Communications Operators

Tests for Listeners. Human beings differ greatly in their ability to make out a message which has been distorted by imperfect equipment and partially drowned out by interfering noise. Tests on random samples from the military population have shown that, under difficult conditions, the best listeners can consistently understand twice as many messages as bad listeners (**9,** pp. 12–14; **5,** pp. 3–8). The results of a typical test are shown in Fig. 2. It is obvious that the random assignment to telephone posts of any personnel, whether well or poorly endowed with this ability, is a hazard to communication.

The ability to listen in noise shows little correlation with general intelligence (**9,** p. 25). Experiments have also shown that this talent cannot be measured by conventional medical tests for acuity of hearing, because (1) against a noisy background a signal cannot be heard unless it exceeds the hearing threshold of all but the most seriously deafened individuals, and because (2) the ability to understand words is a more complex function than that measured by the simple and non-verbal sounds commonly employed in tests for aural acuity (**5,** pp. 14–15).

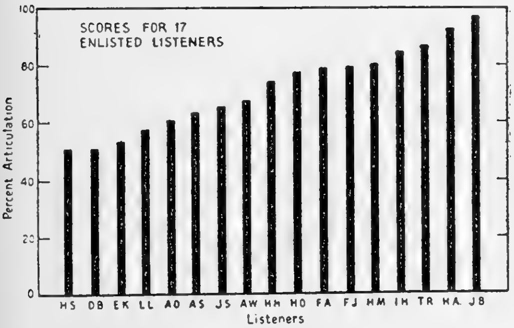

Fig. 2. How listeners differ in their ability to hear words in the presence of noise

Three different tests (*Auditory Tests* Nos. 2, 4, and 8), prepared by the Psycho-Acoustic Laboratory, Harvard University, are available which test this specific ability (**9,** pp. 1–2, and Appendix A). Each consists of speech materials (either single words or sentences) recorded on phonograph discs against a background of masking noise; each may be administered simultaneously, over playback, amplifier, and earphones, to groups of 50 or more men within one-half hour; and each has a reliability of .90 or better. One of these tests (*Auditory Test* No. 8) is of the multiple-choice type, and may be scored on an IBM machine. Available evidence indicates that any one of these tests will yield a valid index to an individual's general ability to understand messages over a communication circuit under conditions made difficult by interfering noise.

Tests for Talkers. Human beings differ even more markedly in their capacity to speak so as to be understood in a noisy environment, than in their capacity to understand what has been said to them. In one group of untrained subjects, for example, speaking under identical conditions, the best talkers were able to make heard five times as many messages as the worst (**10,** IC-60, pp. 6–8; **5,** pp. 8–9). Another group gave the results shown in Fig. 3.

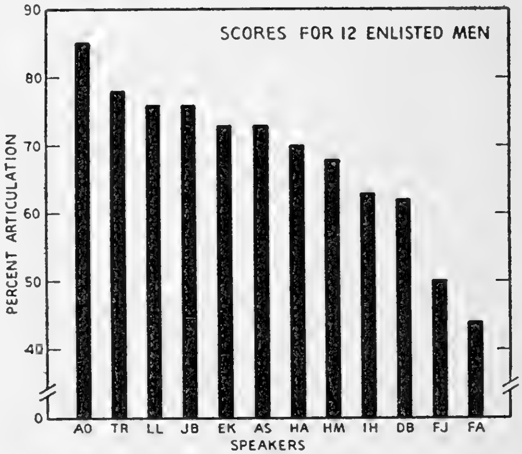

Fig. 3. How greatly a group of talkers can differ in their ability to make words understood by a group of listeners in the presence of noise

Extended experimental analysis indicates that the following attributes of a voice largely determine its audibility in noise (**10,** IC-81, pp. 1–3, 9–11, 27–29; and **20**):

1. The loudness a talker is able to sustain when instructed to speak in the "loudest voice possible without strain."

2. "Consonant strength"—the amount of

speech energy that a talker gets into his consonants. Consonants, rather than vowels, serve as the major clues in identifying a word. At the same time, since they normally contain much less energy than vowels, consonants are the speech sounds most readily masked by interfering noise.

3. "Consonant precision"—the ideal pronunciation for intelligibility in noise exhibits a scrupulous, even exaggerated, articulation of the consonants.

4. "Noise penetrating quality"—subjectively perceived as the possession of a clear, ringing, vibrant, rather than a harsh, muffled, dull voice quality.

Aspects such as the duration of single words, rate of speaking, steadiness of voice-intensity and pace in speaking, the normal pitch of a voice, and the average "shape" of the voice spectrum of an individual talker, were found to be of little or no importance in determining intelligibility. Ratings made by specialists on the basis of a standard college speech interview (depending in considerable part, of course, on aesthetic considerations) were also found to be of very limited validity for the peculiar situation of a military talker who must make himself heard in noise (**10**, IC–67, pp. 3–8).

The ability to talk intelligibly has been found to have little overlap with the ability to listen in noise. Ideally, therefore, personnel for important communications posts should be selected on the basis of both listening tests and talking tests. Talking tests must be administered to one individual at a time, by judges trained to rate a voice according to its possession of the qualities of sustained loudness, precise and powerful articulation of consonants, and a vibrant, penetrating vocal timbre. Reliable talker-rating tests, consisting of a variety of speech materials, and taking from one to three minutes to administer, have been developed for use by the Armed Forces (**12, 13, 14**). The validity of such tests, when administered by experienced personnel, has been established against the criterion of the actual percentage of a man's speech correctly identified over an interphone system by a group of listeners in a noisy environment (**10,** IC–67). A special interphone for rating talkers has also been developed which aids judges in making valid estimates of intelligibility in noise (**11**). This instrument permits simulated engine noise to be introduced into the earphones of speaker and judges during the speech interview, and includes a VU meter by which the loudness of the talker's voice may be measured with precision.

The limited personnel available for assignment to a single submarine, and the need for a telephone talker to possess other than talking and listening skills alone, may make it inadvisable to select operators too rigidly on the basis of these tests. Nevertheless, it would seem wise to have a listening and talking score available for each man in the Submarine Service, so that (1) no man too incompetent will be chosen for a critical communications post, and (2) any submarine lacking a sufficient number of inherently able talkers may recruit additional personnel known in advance to possess the necessary talents.

III. The Training of Communications Operators

Whatever their initial ability, the talking and listening skills of all men can be improved remarkably by practice and by expert instruction. In one experiment, for example, a few minutes of instruction on the proper positioning of the microphone and on the optimal loudness of the voice resulted in an average gain of almost 50 percent in the number of messages a group of speakers was able to make understood against a background of motor noise (**10,** IC–54, and IC–60 pp. 6–12). Skill in listening also improves with practice, although more slowly: in one typical instance, fifteen hours of listening to speech in noise effected a gain of 50 percent in the number of messages one set of listeners was able to identify correctly (**5,** pp. 17–19). Expertly supervised courses in voice communications, especially designed for training military personnel, have been demonstrated to produce important and stable improvements in the general efficiency of telephone talkers and listeners (**15,** p. 6; **16, 17; 23,** pp. 23–40).

A training program, nevertheless, is not in itself an adequate substitute for the selection of telephone talkers on the basis of proficiency tests. For while all talkers, even the poorest, improve with training, the talkers and listeners who were better prior to training tend to maintain their superiority after training as well (**5,** pp. 5–6; **10,** IC-60, p. 12; **15,** p. 17). The optimal program will combine the selection of telephone talkers for their native ability with a course of training to ensure that these abilities are developed to the maximum.

Existing Submarine Talker-Training Program. In 1944, NDRC personnel assisted the Submarine Service in setting up training courses in voice communications at New London, Portsmouth, San Diego, Hunter's Point, Pearl Harbor, and Midway. In its complete form, the course was given in three parts (**24**):

1. A basic course (5 hours) given in a special training room equipped with the various types of submarine communications equipment, as well as with facilities for introducing ambient submarine noise and for recording the voice drills conducted over sound-powered telephone circuits. In this course, all hands were taught general skills in the use of equipment, the art of making oneself intelligible over noise, and the observance of correct voice procedures and circuit discipline.

2. An intermediate course, proceeding as part of the specialized training of torpedomen, diesel engineers, sonar and radar operators, and other groups in the advanced division of the Submarine School. The object was to develop the use of correct voice procedures, phraseology, and enunciation as part of the routine operation of equipment.

3. An advanced course in telephone talking as an integral part of combat-team training, administered after a crew had been assigned to its ship.[2]

The changing circumstances of peace-time training will no doubt impose necessary changes in this voice communications program. However, the mounting difficulty to be expected in maintaining adequate voice communication, in the face of the growing complexity of equipment and procedures and the increasing level of interfering noises, makes it important that the substance of this program be maintained, and even extended. Moreover, since a training course must work against the speech habits of a lifetime, it is to be expected that telephone talkers, once their initial training is ended, will revert to careless speech procedures unless their performance is checked periodically during actual operation of the ship. It would seem desirable, then, that:

1. Ship's officers be indoctrinated with the elements of correct procedures for communicating in noise, and with a knowledge of the proper operation and care of communication equipment, so that they may supervise the observance of these rules by their crews.

2. A submarine talker-training staff be established to make periodic checks on the interior communications of each ship, and to record the flow of messages on the principal telephone circuits. These recordings can then be used as the basis for making corrections, whenever talking performance shall have fallen below the level necessary for the most efficient operation of the ship.

IV. The Standardization of Highly Audible Phraseology

Words, no less than human beings, show extreme differences with respect to the spe-

[2] In addition to special training equipment and training aids, a *Submarine Telephone Talker's Manual* and an *Instructor's Handbook for a Basic Course in Submarine Voice Communications* were prepared and published for use in these courses. The program, methods, equipment, and training aids are described in **24**.

cial requirements of audibility in noise. (See Fig. 4.) For example, a list of words chosen at random was found, under stress of noise, to be correctly identified only half as often as a list of 1000 words selected by special experiment (**7,** p. 60). The success of communications can be aided very materially by the use of words, known to be highly audible in noise, as standard components of all basic messages and procedures. Audibility, though not the sole, is the indispensable attribute of a military vocabulary. If a message is not understood, its other virtues will avail it nothing.

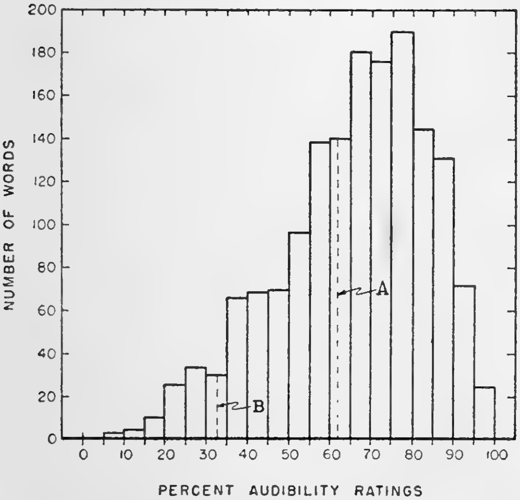

Fig. 4. The distribution of ratings for 1600 words. Each column represents a 5-percent range of audibility. The height of the column shows the number of words falling within each range of audibility. The 1000 most audible words are those which fall to the right of Line A. Line B indicates the average audibility for a separate list of 300 monosyllables and dissyllables chosen at random. This represents the average score to be expected for common words chosen without reference to audibility.

Existing List of Standard Submarine Phraseology. In 1944, the NDRC collaborated with the Submarine Service to develop a handbook standardizing both voice procedures (methods for calling telephone stations and formulating commands, acknowledgments, and reports) and phraseology (specific wordings for all important submarine messages). It was recognized that such standardization would improve the speed and accuracy of voice communications. The talker who has learned a single phrasing for each routine message does not hesitate or fumble for words. The listener who knows in advance that each message is one of a very small number of possibilities has less difficulty in making out what he hears. Men transferred from one ship to another need not waste time nor confuse themselves by learning a new terminology (**24,** pp. 7–12).

Four criteria were used to select a standard submarine phraseology:

1. Conformity to common submarine usage. The final set was chosen from lists of orders and commands submitted by more than twenty submarine commanders.

2. Audibility in noise. Alternative station call-names and other key words and phrases were tested for their relative identifiability over sound-powered phones in the presence of a simulated submarine diesel noise.

3. Freedom from a tendency to confusion. When, in the audibility tests, listeners were found consistently to mistake one message-component for another word of similar phonetic properties, one of these was omitted from the final list. Thus, the word "shut" was chosen over the alternative form, "close," because listeners tended to a dangerous degree to confuse "close" with the other commands, "open" and "blow."

4. Brevity and ease of pronunciation. With other things equal, the simple, succinct message was preferred over a longer or more cumbersome alternative.

The final list of procedures and messages was adopted in a series of conferences attended by a large number of submarine and division commanders. This list was issued and distributed by the Commander, Submarines, Atlantic Fleet, in a handbook entitled *Standard Submarine Phraseology.*

Factors Governing the Audibility of Words. As the construction, the equipment, and the routine operation of submarines change, it will prove necessary to modify and supplement the present list of standard submarine phraseology. To assist in the future selection of new terms which are readily intelligible in noise, it may be useful to summarize the attributes which seem in large part to determine the relative audibility of words (**7,** pp. 74–77; **18, 19**).[3]

1. *The relative power of the component speech sounds.* The most powerful phoneme—"aw" as in all—is 680 times (28 decibels) as intense as the weakest phoneme—"th" as in thin—and the remaining speech sounds are distributed between these extremes (**2,** pp. 70–76). Since the noise of military communications raises the threshold of hearing for sounds other than the noise itself, the weaker sounds are lost first. Under such conditions, it is to be expected that words will tend to be audible in proportion to the power of their component speech sounds, and particularly of their component consonants. On the whole, consonants are the weakest of the speech sounds, yet yield the most important clues for identifying a message.

2. *The spectra of the component speech sounds.* Each speech sound has a characteristic acoustic spectrum, or distribution of energy among the various frequencies (**1; 2,** pp. 28 ff.; **3**). Experiments have shown also that for each sound, certain parts of its characteristic frequencies seem to be the critical regions with respect to its identifiability by a listener (**2,** pp. 279 ff.). Therefore, the relations of the spectrum of a speech sound both to the spectrum of the interfering noise and to the cutoff point and frequency characteristics of a particular intercommunication system will affect its relative identifiability. In addition the identifiability of a speech sound will vary inversely with the number of other speech sounds which it resembles, and for which it is liable to be mistaken.

3. *The number of the component sound elements.* The more sound elements there are in a word, the greater the chance that enough of these will be heard above the noise to identify the word in full. And in general, the longer a word is, the smaller is the number of other words in the language, possessing similar overall phonetic patterns, for which it is liable to be mistaken. Experiment confirms this hypothesis: audibility increases radically with word length, as measured by the number of syllables. For example, in severe noise, disyllables are identified 30 percent more frequently than monosyllables, and trisyllables are identified 50 percent more frequently than monosyllables (**7,** pp. 62, 76; **18,** p. 2).

4. *The competitive context of a word.* The "competitive context" is that set of words of which the listener knows in advance that the stimulus word is a member. The chances that a mutilated portion of a word will be rightly identified tend to vary inversely with the number of other words in the competitive context that it resembles. For example, if the word "power" is partially masked, the relatively low-powered voiceless stop, "p," will be lost first, and the listener will hear only "-ower." If there is no prior limitation of context, the listener must choose between such possibilities as "tower," "cower," "hour," "bower," "sour," "dower," etc. The more a military vocabulary is standardized, the smaller becomes the competitive context of each word in that vocabulary, and the fewer the chances for mutual confusion. Ideally, no two words in a standardized vocabulary should possess a similar overall phonetic pattern.

Rough measurements are available of the relative identifiability in noise of various single speech sounds, and also, of the relative intelligibility of dissyllabic words possessing these alternative stress-patterns: pró-víde;

[3] These factors are in part theoretical, and in part based on experimentation conducted during the last war. The exigencies of war did not permit complete solution of all the problems involved. There is need for a thoroughgoing experimental program to establish in quantitative detail the basic factors governing the audibility of speech in noise, as a foundation for standardizing military oral codes and phraseology.

el-bów; at-tráct; foó-lish (**18,** pp. 4–6, 11–12; **19,** pp. 4–11).

Audibility tests for alternative locutions. On the basis of these various correlates of audibility in noise, an experienced phoneticist could probably set up a relatively satisfactory standard phraseology by inspection alone. The selector may also take advantage of a known list of 1000 highly audible words, chosen from all the familiar words in the English language, for each of which the audibility rating in noise has been established by extensive experimentation (**7,** pp. 57–63, 89–108). In the imperfect state of our present knowledge, however, it is safest to make only a preliminary selection of phraseology on the basis of inspection alone, and to rely on a standard "audibility test" to weed out words which, because of low audibility or tendency to mutual confusion, should have no place in a final list.

The conduct of an audibility test is fully described in **7,** pp. 5–11 and 86–88. A number of speakers read the alternative words and messages to be tested to a group of listeners, over the submarine intercommunication system, against a background of diesel-type noise. The "audibility rating" of a word is the number of times it has been correctly identified, expressed as a percentage of the total opportunities for identification. The available evidence indicates that the audibility of a word is to some extent specific to the response characteristics of the equipment over which it is spoken, the kind of background noise against which it is heard, the local dialects and the educational levels of the talkers and listeners, and the competitive context in which it is presented (**7,** pp. 77–86). For the most precise and valid measurement of the audibility of alternative words and phrases, all these variables ought to be sampled, insofar as they are to be expected in the specific submarine situation. Nonetheless, present indications are that these various factors are not so specific in their effects but that a vocabulary tested over a single interphone, by a fairly representative group of personnel, against one type of diesel noise, will remain highly audible under any foreseeable changes in the conditions of communications on board submarines.

References

A. General

1. Black, J. W. The quality of a spoken vowel. *Arch. Speech*, 1937, **2**.
2. Fletcher, H. *Speech and hearing.* New York: D. van Nostrand, 1929.
3. Laase, L. T. The effect of pitch and intensity on the quality of vowels in speech. *Arch. Speech*, 1937, **2**.
4. Miller, G. A. The masking of speech. *Psychol. Bull.*, 1947, **44**, 105–129.

B. OSRD Reports by the Psycho-Acoustic Laboratory, Harvard University, NDRC

5. OSRD No. 987. *The problem of selecting and training personnel for communication in intense noise.* M. H. Abrams & J. E. Karlin. Nov. 10, 1942.
6. OSRD No. 1571. *Collected informal communications on the basic audibility of English words for use as oral codes, alphabetic equivalents, etc.* M. H. Abrams, J. E. Karlin, J. Miller, F. H. Sanford, & S. S. Stevens. July 9, 1943.
7. OSRD No. 1919. *Vocabularies for military communication in noise.* M. H. Abrams & J. E. Karlin. Aug. 25, 1943.
8. OSRD No. 2038. *Studies on the effect of noise on speech commmunication.* J. P. Egan, J. Miller, M. I. Stein, G. G. Thompson, & T. H. Waterman. Nov. 25, 1943.
9. OSRD No. 3516. *Auditory tests of the ability to hear speech in noise.* J. E. Karlin, M. H. Abrams, F. H. Sanford, & J. F. Curtis. Sept. 1, 1944.
10. OSRD No. 4023. *Speech in noise: a study of the factors determining its intelligibility.* M. H. Abrams, S. J. Goffard, K. D. Kryter, G. A. Miller, J. Miller, & F. H. Sanford. Sept. 1, 1944.
11. IC–71. *A portable interphone for rating and training talkers in noise.* May 1, 1944.

C. OSRD Reports by the Applied Psychology Panel, NDRC

12. OSRD No. 1769. *A speech interview for the selection of telephone talkers.* Report No. 1, Project N–109. Aug., 1943.
13. OSRD No. 1823. *A preliminary study of the*

abilities of rated men to judge speaking performance. Report No. 2, Project N-109. Aug., 1943.

14. OSRD No. 1931. *A study in training classification petty officers to select telephone talkers.* Report No. 3, Project N-109. Nov., 1943.
15. OSRD No. 3178. *Methods of training telephone talkers for increased intelligibility.* Report No. 4, Project N-109. Jan. 28, 1944.
16. OSRD No. 3862. *Training studies in voice communication: I. Can intelligibility of voice communication be increased by training in voice technique?* J. F. Curtis. Report No. 8, Project SC-67. July 5, 1944.
17. OSRD No. 3863. *Experimental investigation of a course for telephone talkers.* Report No. 9, Project N-109. July 4, 1944.
18. OSRD No. 4681. *Phonetic characteristics of words as related to their intelligibility in aircraft-type noise.* Report No. 13, Project SC-67. H. M. Mason. Feb 10, 1945.
19. OSRD No. 5174. *Phonetic characteristics related to intelligibility of words in noise: sounds correctly understood in misunderstood words.* H. M. Mason. Report No. 15, Project SC-67. June 6, 1945.
20. OSRD No. 5413. *Studies of voice factors affecting the intelligibility of voice communication in noise: II. Pitch.* H. M. Mason. Report No. 16, Project SC-67. Aug. 4, 1945.
21. OSRD No. 5461. *Training studies in voice communications: III. Effects of training in articulation.* H. M. Mason. Report No. 18, Project SC-67. Aug. 20, 1945.
22. OSRD No. 5497. *Final report in summary of work on the selection and training of telephone talkers.* Report No. 19, Project N-109. Aug. 27, 1945.
23. OSRD No. 5568. *Final report in summary of work on voice communication.* J. W. Black. Report No. 21, Project SC-67. Sept. 11, 1945.

D. Joint Report by Psycho-Acoustic Laboratory (Harvard University), Applied Psychology Panel (Project N-109), and Columbia University Division of War Research at the U. S. Naval Underwater Sound Laboratory

24. OSRD No. 4795. *A project for standardizing submarine phraseology and developing a training program in submarine voice communications.* M. H. Abrams, J. Miller, L. A. Mallory, G. W. Hibbitt, W. J. Temple, W. D. Neff, & W. L. Jenkins. Feb. 28, 1945.

Chapter 11

VOICE COMMUNICATIONS: EFFECTS OF MASKING AND DISTORTION

GEORGE A. MILLER

Harvard University

Carefully selected and well trained personnel may use a highly audible vocabulary and adhere to standard procedures, but still be helpless victims of inadequate equipment. If the signal arrives weak and distorted while noise roars into the listener's ears, vocal communications will be severely crippled, perhaps impossible.

The principal characteristics of the equipment which determine its performance are: (1) the signal level, (2) the noise level, (3) the frequency-response characteristic, and (4) the type and degree of non-linear amplification. Articulation tests (**8, 11, 14**) have been used to evaluate the importance of these characteristics for the intelligibility of speech. We will review briefly some of the results of this work, point out occasional applications to problems of military communication, and indicate how the subjective methods of evaluation might be predicted from a knowledge of the physical characteristics of the system.

I. Relation between Intelligibility and Intensity of Speech

If the listener and talker are connected by an electrical communication system, it is a simple matter to explore the relation of intelligibility to intensity by varying the amount of amplification or attenuation in the system. Over a wide range of intensities listeners will respond quite uniformly by writing down words or sentences spoken by an announcer. At very low intensities some of the speech sounds become inaudible and some of the test items are missed. At very high intensities the sound becomes painful and the listener refuses to cooperate. The determination of these limits tells us the range of speech intensities that is available for vocal communications.

The most intense vowel sound, the vowel in *ought*, is just detectable half the time when the average level of the speech is zero db *re* 0.0002 dyne per cm.^{2}, but the faintest sound, the *th* in *thin*, does not become audible until this average level is raised 25 or 30 db (**10**). The presence of about half the words can be detected when the average level of speech in a free field is about 5 db (**17**). When the speech is heard via earphones, the threshold of detectability for half the words may be about 5 db higher (**16**).

The intensity necessary to identify half the test items correctly depends upon the type of speech material. Nonsense syllables and monosyllabic words give 50 percent articulation scores at levels between 30 and 35 db (**6, 12, 17**). Polysyllabic words, sentences, and continuous discourse give thresholds in the range from 20 to 30 db (**6, 16, 25**). Binaural listening gives thresholds of intelligibility about 3 db lower than monaural listening (**25**).

At the other end of the scale, listening becomes uncomfortable at about 117 db, begins to tickle the ear at 128 db, and is painful at 138 db. After some experience with these intense levels, however, thresholds of discomfort, tickle and pain are raised about 10 db (**6**).

The usable range of speech intensities is, therefore, well over 100 db, a power ratio of 10^{10} to 1.

II. The Masking of Speech

The effect of an interfering noise is to raise the thresholds of detectability and intelligibility, and so to decrease the range of intensities available for vocal communication. This shift in the threshold is the measure of *masking*. Military communications are chronically afflicted with noise, and an estimate of how much noise can be tolerated in any given situation is a highly practical undertaking.

Pure tones are relatively ineffective in masking speech. If they are intense enough, low-frequency tones can mask the entire range of speech frequencies, but high-frequency tones mask only a restricted range of frequencies. The listener is able to decipher the speech on the basis of the unmasked portion of the speech spectrum (**27**). Complex tones of low frequency are more disrupting, but the greatest danger comes from continuous spectrum noise.

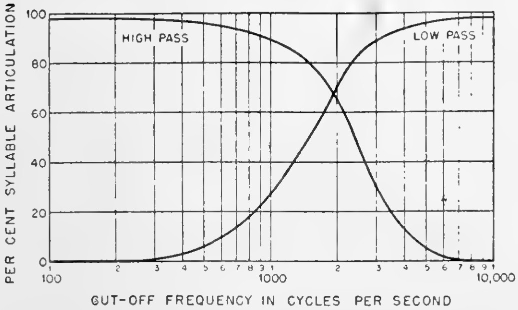

Fig. 1. The relation between articulation score and the cut-off frequency of high-pass and low-pass filters. Composite for men's and women's voices. (After French and Steinberg, **12**)

The simplest type of noise to specify for experimental purposes is one which has a uniform level per cycle over the range of speech frequencies (100–8000 cps). This type of noise is sometimes called *white noise.* When the sound pressure level of this noise is 40 db, the threshold is shifted about 12 db. For levels above 40 db, this noise produces a one db shift in the threshold for every one db increase in the intensity of the noise (**16**). This relation is sometimes expressed by saying that the signal-to-noise ratio necessary for communication is constant over a wide range of intensities. Only at very low and very high noise levels must the signal-to-noise ratio be increased. Experience has shown that for most continuous spectrum noises encountered in military situations the signal-to-noise ratio should exceed +6 db for satisfactory communications, although the presence of speech is usually detectable at ratios as low as −18 db.

Aspects of the noise which are known to influence its masking effectiveness include its spectrum (**22**), its temporal continuity (**22**), the phase relations between the ears (**20**), and the meaning of the noise to the listener. Further research is needed to assess the importance of these variables.

III. Frequency Selectivity

Voice communication equipments were markedly improved during World War II, especially as regards the reduction of frequency distortion. A great deal of frequency distortion can be tolerated under otherwise ideal conditions, but a system which passes a narrow range of frequencies quickly shows its inadequacy when tested under the stress of noise.

In laboratory tests of frequency selectivity it is desirable to use filters which pass uniformly a given range of frequencies and greatly attenuate all components of the signal which lie outside this range. Although such systems are rarely encountered in practice, they have the advantage of enabling us to estimate the relative importance of the various frequencies in the spectrum of human speech.

The results of articulation tests conducted in the quiet with nonsense syllables distorted by high-pass and low-pass filters are summarized in Fig. 1 (**12**). If a high-pass filter is used to select only components above 6000 cps, no syllables will be correctly recorded by the listeners. When only frequencies above 3000 cps are passed, 30 percent are discriminated. Frequencies of 1000 cps and above give 90 percent syllable articulation. The curve for low-pass filters can be similarly interpreted.

The two curves of Fig. 1 cross at 1900 cps. This value is obtained for men's and women's voices. For men's voices alone the crossing point is at 1660 cps (**1**). The frequencies above this crossing point contribute as much to intelligibility as do the frequencies below. If we define the range from 200 to 6000 cps as contributing 1.00 to intelligibility, then 200 to 1900 cps contributes 0.50, and 1900 to 6000 cps contributes 0.50, and the range of speech frequencies is divided into two bands of equal importance. It is possible to repeat this process of subdivision until the frequency limits of x bands, each contributing $1/x$ to intelligibility, are determined.

This weighting of the frequency scale according to importance for intelligibility enables us to compare communication systems with different frequency-response characteristics. It is necessary, however, to weight the contribution of each of the x bands according to the signal-to-noise ratio in the band. A band begins to contribute to intelligibility when it is only slightly above the threshold of detectability, but does not make its maximum contribution until it is 50 db above the threshold.

It is possible, by weighting the frequency and intensity scales in this manner, to predict the adequacy of a communication system from a knowledge of its frequency-response and amplification, the noise level, and the type of speech materials (**1, 2, 3, 4, 5, 12 24**). Such computational procedures are a great convenience in the design of communication systems, and make it possible to minimize the amount of articulation testing required. This does not mean, however, that articulation testing methods can be abandoned. The computations provide a first approximation, but their accuracy must be increased. Many of the things which can go wrong in a communication system are not considered. Much further research is needed before we can be satisfied with this theoretical interpretation of the perception of speech. An effort should be made to maintain the scientists and facilities for such experimentation.

Sound-powered telephones provide an example of the practical applications of this research. The output level of a sound-powered system can be increased if the frequency range passed by the system is reduced. Is it better to have a wide-band system with a weak output, or a narrow-band system with a loud but distorted output? The information necessary to answer this question is summarized by the equal-articulation contours in Fig. 2 (**9**). These curves were determined with a variety of band-pass systems

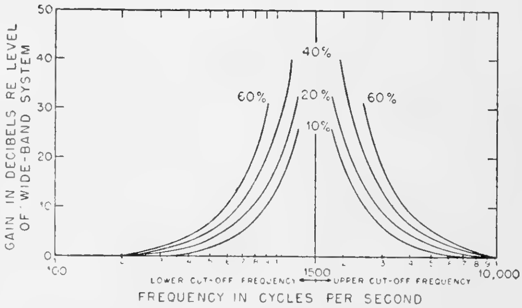

Fig. 2. Equal-articulation contours for band-pass systems in noise. The parameter is the articulation score. The abscissa indicates the lower and upper cut-off frequencies of the band. The ordinate gives the increased amplification necessary for the narrow-band system to maintain the same articulation score as a wide-band system. Men's voices. (After Egan and Wiener, **9**)

tested with nonsense syllables in the presence of uniform white noise. On the ordinate, zero db is the intensity of the speech in a wide-band system. In order to maintain any given articulation score, the level of the speech in a narrower band must be increased. The curves are used in the following way. Assume that a band-pass system extending from 950 to 2400 cps is to be widened to 550 to 4200 cps. In order to maintain an articulation of 10 percent, how much loss in level can be tolerated in the wider band? Fig. 2 indicates that the articulation score of 10 percent could be maintained even though the level of the received speech provided by the wider band was 9 db less than that provided by the narrower band. Now, suppose that the signal-to-noise level at which both systems give 10 percent articulation can be improved 5 db. This improves the narrow system from 10 to 20 percent, but improves the wider system from 10 to 50 percent articulation.

This family of curves is constructed with 1500 cps as the geometric center frequency of the band, but to a reasonable approximation the same contours can be used for other center frequencies by shifting the entire family horizontally one way or the other on the semi-log plot.

IV. Amplitude Selectivity

Amplitude, or non-linear, distortion occurs whenever a signal is passed through a circuit which transmits some amplitudes of the wave and not others. The most frequent type of amplitude selectivity is due to circuits which transmit low-amplitude parts of the wave more efficiently than high-amplitude parts. Reduced to its simplest form, this type of distortion is called peak clipping—the peak voltage swings are clipped symmetrically off the top and bottom of the speech wave.

Although peak clipping degrades the quality and naturalness of speech, articulation tests indicate that a surprisingly large fraction of the speech wave can be eliminated before intelligibility is affected (**10, 13, 19, 28**). The resistance to the effects of peak clipping is shown graphically in Fig. 3. More than 95 percent of monosyllabic words are recorded correctly even after 24 db of peak clipping, when all that remains of the speech wave is the $\frac{1}{16}$ nearest the center axis. About 70 percent of the words are understood after infinite peak clipping which reduces speech to a succession of rectangular waves (**21**).

If the system passes the high-amplitude parts of the wave and rejects the low-amplitude parts, it is a case of center clipping. The center of the wave nearest the time axis is clipped out and thrown away. With it goes the intelligibility of speech, as is evident from Fig. 3. Clipping out only one-quarter of the speech wave (2.5 db center clipping) reduces word articulation to 30 percent. This type of distortion results in sounds that resemble static more than speech. The disastrous effect of center clipping is due principally to the elimination of the faint consonant sounds which provide most of the intelligibility. Peak clipping, on the other hand, tends to emphasize the consonants relative to the vowels.

Some types of amplitude selectivity are intolerable, while other types can be accepted and even used to practical advantage. Peak clipping circuits have actually been recommended in some situations—e.g., in amplitude-modulated radio transmitters (**15, 18, 23, 26**), radio receivers (**19**), and hearing aids (**7**).

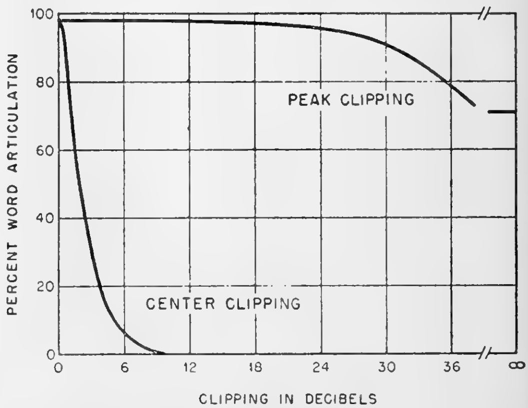

Fig. 3. The effects upon articulation scores of clipping the peaks versus clipping the center out of the speech wave

V. Evaluation

We have touched only the high spots in this outline of the effects of masking and distortion upon voice communications. Many other masking signals and types of distortion have been studied and many remain for future study. This work should, and probably will be, continued. Every new communication device brings novel masking sounds, new possibilities for distortion. It would be a great pity if the research effort of the Navy were to be restricted to the specific problem posed by this or that engineering innovation. During a war this orientation toward testing individual pieces of equipment is necessary and unavoidable. Now, with the heavy weather past, we may gain a glimpse of the sun to take our latitude and replot our true course.

The problem for future research is to contruct a general theory of auditory perception that will bridge the gap between the perception of meaningless tonal stimuli and the perception of meaningful spoken messages. The present state of the art indicates that such a theory is possible to achieve, but it has not been achieved as yet. More information is needed. This information will come in part from sensory physiology, in part from phychophysical studies of the perception of tones, and in part from articulation testing procedures. We need to know more about the perception of the individual speech sounds, synthetic speech, complex tonal patterns and noise, and the temporal aspects of sound stimulation. As this information accumulates we should be able to draw a clearer picture of the terminal component in every voice communication system—the listener.

References

1. Beranek, L. L. The design of speech communication systems. *Proc. Inst. Radio Engin.*, 1947, **35**, 880–890.
2. Collard, J. A theoretical study of the articulation and intelligibility of a telephone circuit. *Elec. Comm.*, 1929, **7**, 168–186.
3. Collard, J. Calculation of the articulation of a telephone circuit from the circuit constants. *Elec. Comm.*, 1930, **8**, 141–163.
4. Collard, J. A new criterion of circuit performance. *Elec. Comm.*, 1933, **11**, 226–233.
5. Collard, J. The practical application of the new unit of circuit performance. *Elec. Comm.*, 1934, **12**, 270–275.
6. Davis, H. (ed.) *Hearing and deafness.* New York: Murray Hill, 1947.
7. Davis, H., Hudgins, C. V., Marquis, R. J., Nichols, R. H., Ross, D. A., & Stevens, S. S. The selection of hearing aids. *Laryngoscope*, 1946, **56**, 85–115, 135–153.
8. Egan, J. P. *Articulation testing methods II.* OSRD Report No. 3802 (Psycho-Acoustic Laboratory, Harvard University, Aug. 25, 1943). Available through the Office of Technical Services, U. S. Dept. of Commerce, Washington, D. C. PB No. 22848.
9. Egan, J. P., & Wiener, F. M. On the intelligibility of bands of speech in noise. *J. acoust. Soc. Amer.*, 1946, **18**, 435–551.
10. Fletcher, H. *Speech and hearing.* New York: D. Van Nostrand, 1929.
11. Fletcher, H., & Steinberg, J. C. Articulation testing methods. *Bell Syst. tech. J.*, 1929, **8**, 806–854.
12. French, N. R., & Steinberg, J. C. Factors governing the intelligibility of speech sounds. *J. acoust. Soc. Amer.*, 1947, **19**, 90–119.
13. Gemelli, A. Observation sur le phonème au point de vue de la psychologie. *Acta Psychol.*, Hague, 1938, **4**, 83–112.
14. Goffard, S. J., & Egan, J. P. *Procedures for measuring the intelligibility of speech: sound-powered telephone systems.* Report PNR-33 from the psycho-Acoustic Laboratory, Harvard Univ., Feb. 1, 1947.
15. Gross, N. B., & Licklider, J. C. R. *The effects of tilting and clipping upon intelligibility of speech.* Report PNR-11 from the Psycho-Acoustic Laboratory, Harvard Univ., April 15, 1946.
16. Hawkins, J. E. *The masking of signals by noise.* OSRD Report No. 5387 (Psycho-Acoustic Laboratory, Harvard Univ., Oct. 1, 1945). Available through the Office of Technical Services, U. S. Dept. of Commerce, Washington, D. C. PB L68916.
17. Kryter, K. D. Effects of ear protective devices on the intelligibility of speech in noise. *J. acoust. Soc. Amer.*, 1946, **18**, 413–417.
18. Kryter, K. D., Licklider, J. C. R., & Stevens, S. S. Premodulation clipping in AM voice communication. *J. acoust. Soc. Amer.*, 1947, **19**, 125–131.
19. Licklider, J. C. R. Effects of amplitude distortion upon the intelligibility of speech. *J. acoust. Soc. Amer.*, 1946, **18**, 429–434.
20. Licklider, J. C. R. The influence of interaural phase relations upon the masking of speech by white noise. *J. acoust. Soc. Amer.*, 1948, **20**, 150–159.
21. Licklider, J. C. R., Bindra, D., & Pollack, I. The intelligibility of rectangular speechwaves. *Amer. J. Psychol.*, 1948, **61**, 1–20.

22. Miller, G. A. The masking of speech. *Psychol. Bull.*, 1947, **44**, 105–129.

23. Miller, G. A., & Mitchell, S. Effects of distortion on the intelligibility of speech at high altitudes. *J. acoust. Soc. Amer.*, 1947, **19**, 120–125.

24. Pocock, L. C. The calculation of articulation for effective rating of telephone circuits. *Elec. Comm.*, 1939, **18**, 120–132.

25. Shaw, W. A., Newman, E. B., & Hirsh, I. J. The difference between monaural and binaural thresholds. *J. exp. Psychol.*, 1947 **37**, 229–242.

26. Smith, W. W. Premodulation speech clipping and filtering. *QST*, 1946, **30**, 46–50.

27. Stevens, S. S., Miller, J., & Truscott, I. The masking of speech by sine waves, square waves, and regular and modulated pulses. *J. acoust. Soc. Amer.*, 1946, **18**, 418–424.

28. Strecher, F. Verständlichkeit und Lautstärke bei Frequenz- und Amplitudenbegrenzung. *Akust. Z.*, 1936, **1**, 174–178.

Chapter 12

VISUAL COMMUNICATION

WILLIAM S. VERPLANCK[1]

Indiana University

Foreword

It is not possible in this chapter to survey the whole field of visual communication, ranging as it does from the austere and complex profundities of the mathematical journals to the questioning look of a sailor meeting a girl. In man's history, visual communication has played no small role, even if writing is excluded from consideration. In some cultures, visual communication has assumed the function of spoken language, whether supplementary to the spoken language, as in the Italian's use of gesture, or in place of it, as in the sign language of the Plains Indians. We must here be concerned with the application of visual communication systems to the modern submarine, and more particularly with their possible usefulness as a supplement to, or a substitute for, voice communication, where the latter has been solely employed and found not entirely satisfactory. We shall, then, attempt to clarify some of the problems of visual communication, to indicate some areas where visual communication might well be employed, and to survey in a summary fashion the areas of psychological research which may be expected to yield the information which is needed for the exploitation of visual communication to its fullest.

Introduction

Military commands, whether submarines, aircraft, or regiments, must act as units. The activities of individual machines and men must be coordinated so that the unit can function with fullest efficiency, both in the appraisal by command of the situation in which it is placed, and in the execution of maneuvers dictated by command's evaluation of and response to that situation. Individual men must be trained and individual machines designed so that both fit into a single organization capable of immediate and precise action. The means by which such integration of the behavior of single units is effected has been called "coupling," after the terminology applied to purely mechanical systems, and may be achieved relatively readily when we are dealing with machines alone. When, however, the human individual, enters into a couple with a machine, or with another individual, many and complex difficulties may arise, and the behavioral properties of the human must be carefully considered. The greater difficulties and frequently lessened reliability of systems in which humans enter as couples have led to a regular and consistent tendency to eliminate the human from couples wherever possible—to move toward the push-button war.

We are not, as yet, able to design and construct such purely mechanical combat units; there remains the very live problem of the production of more efficient, rapid, and accurate couples involving the human. Research has been directed to each of the three types of couples; machine to man, as in instrument and radar design and display; man to machine, as in investigations of instrument controls, and man to man, as in communication.

The control of the behavior of one individual by the behavior of another individual is

[1] The author wishes to express his thanks to those who have offered helpful comments regarding the problems of visual communication. Particularly, he wishes to thank the officers and men with whom he consulted at New London and Key West.

the problem of coupling with which we are concerned. Such control is effected by the stimulation of the sensory mechanisms of the second individual (the receiver) by the behavior of the first (the sender), in such a way that he will respond in a predictable fashion. The problem of communication is the problem of the effectuation of such stimulation, usually across a distance which is not negligible.

Terms

In accordance with this analysis, a *message*, or *information*, is that complex of symbols produced by the behavior of the sender to which the receiver is expected to respond in a predictable way, according to the signification of the symbols. The *sender* transmits the *message* to the *receiver* through some *communication system*. In many instances, it is desirable for the receiver to return an *acknowledgment* to the sender, verifying in detail the receipt of the message. This acknowledgment is often a transmission of the same message back to the sender by the receiver.

Communication systems may be considered as *direct* or *mediate*. *Direct systems* are those in which the sender is in direct contact with the receiver through the physical medium of communication and transmits the message by voice, by gesture, or other previously agreed upon code. In *supplemented direct systems*, the sender or the receiver, or both, may make use of special aids to communication through the medium. For the sender, these may include signal lights, semaphore flags, megaphones, or similar devices for increasing the distance across which the symbols produced may be seen. For the receiver, special aids to vision or hearing may be employed, such as binoculars, snooperscopes, microphones and amplifiers, and so on.

Mediate system is a term reserved for those methods of communication which do not require the possibility of direct stimulation of the receiver by the sender, but which make use of more or less elaborate mechanical or electronic devices for the conversion of the energies controlled by the behavior of the sender into other forms of energy, which ultimately stimulate the receiver. Such systems range in complexity from the handwritten message to television, and include sound-recording methods and radio. It is apparent that both direct and mediate systems may involve either the receiver's vision or hearing, or both.

Two general classes of messages may also be distinguished. The first we may call *persistent*. Persistent communications are those which are not necessarily addressed to a specific receiver, which neither require nor expect immediate action of the recipient, and which are not needed by the recipient as a basis for urgent action. They are typically recorded in some manner, so that they may be referred to or acted upon at a later date. Such messages include statements of basic procedures, general instructions, and so forth.

In contrast, *transient communications* are those which either require immediate and precise response on the part of the receiver, or the contents of which are urgently needed by the receiver as the basis for immediate decisions on a course of action. Transient messages include combat information, commands, and status information. They invariably require extremely rapid transmission, and optimally, it appears, should be equally rapidly acknowledged in detail. They are seldom recorded, except for historical purposes.

Two media of communication have been of military importance: light and sound. Light, as a medium of communication, involves the production of signals which will ultimately stimulate the visual apparatus of the receiver, and sound involves the production of stimuli affecting his auditory system. These media define *visual* and *auditory communication*, respectively. In connection with communication, we shall employ the term display, which refers to the set of signs produced in the visual or auditory field of the receiver by the sender, and to which the receiver responds.

Persistent communications have tended to be largely visual, owing to the multiplicity of such symbols which are available in our culture, and to the relative ease with which they may be recorded and reproduced. Transient communication, on the other hand, tends to be auditory, and more specifically, usually employs the voice, since it requires both speed and well-established habits of response to symbols. Speech meets these requirements well. It is the area of transient communications in which we shall examine the possible role of visual methods, since their utility there has been but superficially investigated.

A final type of distinction which it will be convenient to make is that between *exterior communications* and *interior communications*. The former refers to couples between persons on a vessel with persons off of it, and the latter refers to communication among persons all of whom are on one vessel, but who may be in different compartments.

Just so long as information transmitted by one person to another is not invariably received correctly, and at the appropriate time, and acted upon appropriately, there exists a need for continued research on communication. Since developments to date have by no means eliminated a variety of difficulties in communications, it would appear that a field for research remains.

Properties Desirable in Transient Communication Systems

Before discussion of the several visual systems which have been employed for transient communication, it is advisable to survey briefly the considerations which must be taken into account in designing and evaluating such systems.

Speed

A communication system carrying transient messages must transmit them without delay to the receiver, and must stimulate the receiver in such a way that he, in turn, does not delay in responding. Systems, then, which require a time-consuming procedure of transmission, such as semaphore or manual radio code, or of reception, such as blinker, are inferior to such systems as radio teletype, telephone, and the engine-order telegraph.

Precision

A communication system should be specifically designed to carry stated information in a precise and unequivocal way. It should permit no misunderstanding of the message. Signs employed should be unequivocal; as stimuli they should be easily discriminable. The diving alarm is an example of such a precise and unequivocal auditory communication system, as are the engine-order telegraph and the sound-powered telephone when used in the absence of background noise by well-trained personnel.

Identification and Acknowledgment

The system should provide for the clear identification of the sender, and of the intended receiver. It should further provide for a clearcut acknowledgment of the message, which is preferably a repetition of the message in the reverse direction, since this procedure ensures that misunderstandings are immediately revealed.

Simplicity

Simplicity is a merit in communication systems. This simplicity should lie not only in the design and construction of the instruments employed, but also in the method of use. It should be possible for persons of inferior intelligence and training to use them properly.

Attensity

The system should demand the attention of the receiver. Since transient communications are almost always aperiodic, the sender must be able to prepare or set the receiver for the message which is to be transmitted. The telephone bell is an effective warning signal, and in general, sound systems more readily lend themselves to such signals, since they do not require the receiver to be physically oriented towards a particular spot, as does a purely visual communication display. In exterior communications, the watch is responsible for picking up the signals of ensuing messages. It is doubtful if sufficient work has been done in investigating the usefulness of bright flashing lights for this purpose in interior communications.

Dependability

It is obvious that any system of communication should be ready for efficient use at any time. Lack of dependability is one of several factors which have led to the general distaste for direct exterior visual communication, so that it has become but rarely employed by submarines. It is this factor also, which is at the basis of criticism of certain of the auditory systems.

Security

Although security is not an important factor in interior communication, it is a prime problem in exterior communication. Direct visual communication is easily intercepted, and even though the exact content of a message may not be intelligible to an interceptor, nevertheless, the detection of such a message reveals not only the presence of the sender, and of the receiver, but it also may lead to accurate inferences with respect to their intentions. Many and ingenious devices, as well as codes (e.g., the infra-red or ultra-violet light telephone of the Japanese and Germans and our infra-red blinker (**8**)), have been developed to provide secure external communication systems. The need for security is another of the factors which have almost eliminated exterior visual communication from the submarine service.

Freedom from Hostile Jamming

The activities of the enemy should not make communication over the system impossible. The use of the radio transmission of noise by the enemy to make radio voice or code unintelligible, and the use of smoke screens to make visual communication impossible, can be most effective in reducing combat efficiency when it is dependent on exterior communications. An optimal exterior communication system is not easily subject to such interruption.

Freedom from Interruption

A communication system should be designed so that it will transmit only those messages which it is supposed to transmit. The system should be easily limited to use for its specified purposes, so that no problem of circuit discipline need arise. In practice, a serious objection to many types of auditory communication systems is the ease with which they lend themselves to casual, unnecessary, and trivial communications which may render them temporarily inaccessible for urgent messages.

Planning

A communication system must be carefully planned, after analysis of its intended functions, with respect both to the locations of the sender and receiver who communicate over it, and to the number and variety of messages which must be transmitted over it. Careful planning of the communication system on the basis of its operational function is essential for efficiency and freedom from trouble. Such analysis not only makes it possible to reduce the number of individuals who must make use of the system (and so to free it from an undue amount of traffic, with consequent jamming at times of stress), but also to reduce the number and complexity of the messages which it must carry. When the analysis of operational demand is complete, the communication system may be designed to be just as versatile in the variety of messages it will carry as it need be, and no more. The fewer messages and the greater the simplicity of content which it must carry, the easier, more precise, and more rapid can be the receiver's response, since the number of signs to which he must respond are few, and do not require complex discrimination.

Demand on the Senders and Receivers

Here we face the basic psychological problems of communication: the control with certainty of the response of the receiver. Many of the difficulties found in current communication systems stem from the demands placed upon the receiver by the complexity of the messages transmitted, and by the speed with which they must be responded to. With highly selected and trained personnel, the problem is a serious one; as selection and training procedures are less effective, the less effective the complex communication system and the more difficult the problem. The *span of apprehension* and the *immediate memory span* of the receiver must both be carefully considered.

The span of apprehension has been perhaps most effectively defined and studied in the field of vision. Essentially, it is the number of discrete stimuli which, presented simultaneously for a brief period, can be accurately responded to. Many recent investigations have clarified considerably the controlling variables, which prove to be the length of time through which the stimuli are presented to the subject, and secondly (for stimuli of very short duration), the brightness of the stimulating field. The number of stimuli which can be simultaneously accurately responded to is sharply limited for very short exposures; it is approximately four at $\frac{1}{10}$ second (**4**). This process of immediate discrimination of numerosity has been termed "subitizing" by Reese (**12**). When the time increases up to tens of seconds, the number of stimuli correctly responded to increases regularly, as the subject beings to count, whether by single stimuli, or by groups of them (**12**). Such discrimination of number is a special case of the fundamental case of how complex a stimulus field the organism can respond to appropriately and accurately in a brief time. It may be presumed that the same variables, time and intensity, play much the same roles when stimulus dimensions other than numerosity are concerned in the discrimination. The equivalent problem in hearing has not been as well defined, although there is unquestionably a problem of discrimination (i.e., comprehension) for rapidly delivered speech. On the other hand, it is well established that where discrimination is difficult, whether because of speed or lack of clarity of delivery, or because of background noise at or above the loudness of speech, the receiver will respond inappropriately (**7, 10**). The response to the message becomes independent of the actual message and depends upon the receiver's dominating motivation at the time of misapprehension and on his own working vocabulary; unfamiliar words in the message are responded to as familiar words which sound somewhat like them.

The second span with which we must deal in communication is the *immediate memory span*. This is defined as the maximum length of a series of stimuli which can be repeated correctly by the receiver after one presentation of the series. Thus, the number of integers in the longest series which the individual can repeat correctly immediately after he has heard it defines the immediate auditory memory span for integers. Two different such "spans" may be identified in vision: the maximum size of a set of different stimuli which the subject can report correctly after having seen the variegated set when it is presented either (a) simultaneously, or (b) successively. The latter is obviously closer to the auditory memory span, and the former to the visual span of apprehension.

The immediate auditory memory span is one of the performances which define the scientific concept of intelligence. Systems which place undue demands on these psychological functions of the receiver may be expected to produce unsatisfactory and unreliable communication. "The men don't understand what is said," and, "They forget what they heard," are not uncommon complaints about voice systems. When voice systems carry only a few well defined messages whose content has been carefully selected for intelligibility against a noise background, for precision, and for the avoidance of sounds which are similar to one another, communication becomes quicker and more accurate, and the difficulties placed upon the system by the span of apprehension and of immediate memory tend to evaporate.

Minimal Need for Training

In general, owing to our well-established speech habits, auditory systems require far less training of both sender and receiver than do visual systems. Voice requires a minimum of training. Systems which do not utilize speech, and which are extremely versatile in the variety of messages which they can carry require relatively long training, whether auditory or visual. The studies of Keller and his associates (**5**) on radio code transmission during the recent war showed that such training involves two complex processes.

1. *Differentiation of response.* The sender must learn to give accurately and rapidly a variety of extremely complex sequences of responses, which are sharply differentiated from each other and from any other similar responses he might make. The radio code sender must learn to transmit words by a complicated and specific pattern of muscular activities, which have the effect of closing an electrical circuit to produce a temporal patterning of electrical impulses.

2. The receiver, on the other hand, must develop *a sensory discrimination.* He must learn to respond differentially among a great variety of complex stimuli which closely resemble one another.

Both processes are basic to the development of versatile communication systems which do not use speech. The necessity of establishing differentiation among responses and discriminations among stimuli, again emphasizes the need for necessary messages to be analyzed with the view of reducing the number of symbols which must be employed in a communication system. As each new response of the sender is added, the difficulty of establishing differentiation and discrimination increases; the time required for training is an exponential function of the number of different discriminations and differentiations which must be made.

This suggests strongly that if modifications in naval communication systems are made, they should make use of already established habits of discrimination and response, so far as possible. Systems which represent further developments of already existing systems should be sought.

With these criteria, it becomes possible to enumerate some of the difficulties of transient auditory communications that suggest that the feasibility of visual systems be investigated as possible complements or substitutes for existing auditory systems.

Some Remarks on Systems

Two limitations of auditory communication systems stand out as most significant. The first of these is the presence, in many situations in which accurate communication is necessary, of masking noise at intensity levels so high that the individual is not able to discriminate the message from the masking background noise against which it appears. In submarines, the ambient noise may reach and remain at levels of 80[2] or 100 decibels (**6**).

The second limitation is that auditory systems depend greatly for their efficiency upon the intelligence and training of the receiver. He must comprehend the message, and retain it, prior to his response. If information or commands are received but misunderstood, or if they are forgotten before action has been taken, unfortunate consequences may ensue. The "Aye, aye" or "Very good" of a simple acknowledgement may obscure failure to understand or remember. Auditory communication, then, places great demands not only on auditory

[2] Harris, J. D. Personal communication.

acuity, but also upon the capacity of the individual to comprehend and to retain, i.e., upon the spans we have discussed above.

Problems in auditory communications do not cease with these major ones, however. Poor intelligibility of electrically transmitted speech (without respect to ambient noise) may lead to misunderstanding. Another grows from American habits of use of the telephone. Circuits which may be called upon for instantaneous use may be cluttered with idle chatter. The problem of circuit discipline was nearing solution toward the end of the recent war, but was not universally solved. Telephone circuits are often crowded circuits.

Other difficulties arise in special circumstances. An instance has been reported in which the sounds of loud auditory communications proved objectionable, if not demoralizing, to members of the crew of a submarine after a severe depth-charging. Doubtless there were other such cases.

These difficulties may appear as relatively minor. They must, however, be recognized and, if possible, circumvented. The rapidity of auditory communication, and the simplicity of the apparatus required, ensure that audition will remain the normal and most important method of transient communication. Nevertheless, visual communication may prove to be a useful supplement to voice in reducing the hazards created by the difficulties we have noted, since it is in those areas where auditory communication is weak that visual communication may possess special advantages. It must be recognized, however, that these advantages are not yet fully realizable.

Visual systems have remained largely undeveloped for transient communication, whether interior or exterior. The development of the telegraph, telephone, and radio has aborted their development. Visual communications are still, to some extent, employed in ship to ship communication; the signal flag and blinker remain. Even here, however, they have been found useful only in the absence of efficient auditory systems, where distance or inconvenience makes direct communication by TBS or code inconvenient or impossible. Recent developments in electronics have served still further to reduce the role of vision in communication. During World War II, intensive research was performed on auditory communication—on the intelligibility of sounds, on microphones and other equipment, on telephone circuits, on radio telegraphy, and on code signalling. However, practically none was done on the still primitive visual systems, such as blinker-signalling and semaphore, except to obtain data on visibility ranges. Neglect of the possibilities of visual systems for interior communications is also evident.

Visual communication would seem, potentially or actually, to have the following advantages. First, it is fully resistant to masking by noise, and is easily rendered equally resistant to interference by other type of competition. Second, the visual signals employed may be made to persist through a period of time great enough so that the immediate memory span and the span of apprehension of the receiver are not exceeded, as often happens in hearing. Third, a wide variety of information may be simultaneously transmitted. Fourth, several individuals may simultaneously receive the message, which is impossible with the telephone. Fifth, more rapid, more accurate methods of acknowledgement are made possible. Sixth, visual communication designed for specific purposes is not apt to be subject to circuit jamming.

The difficulties of visual communication are not to be overlooked, and stem primarily from their lack of flexibility, that is, from the limitation of the number and variety of the messages which may be transmitted without the existence of some complex and bulky system, such as the teletype, and without extremely long training. A completely flexible visual communication system that does not utilize the printed word would require such a tremendously complex series of visual symbols, each with its own meaning, that only with extreme difficulty could sender and receiver learn them. Planning and content analysis of communications, directed at reducing the symbols necessary to a minimum, is a necessary starting point for the further development of visual communication.

Variations in Visual Stimuli which may Serve as the Basis of Communication Systems

Before examining the present status or the potentialities of various types of visual communication systems, let us inquire into some of the stimulus variables which might serve as symbols in the transmission of messages. Since these dimensions are treated in detail elsewhere in this book, they will be considered only for their probable utility in communication. Except where otherwise specified, the discussion will relate to the properties of a *single* visual stimulus.

Brightness

The human being is capable of making numerous discriminations with respect to the brightness of a visual stimulus. However, such discriminations are unreliable, since they occur in the absence of a uniformly illuminated background, with which the stimulus can be compared, and depend greatly upon the level of adaptation of the eye. Brightness is, moreover, a significant variable only at night, when the background illumination is negligible. In this situation, it is a remarkably untrustworthy basis for establishing discriminations, since it will depend, not only on the brightness of the stimulating source, but also on the distance of the source from the receiver, and on the atmospheric attenuation of the light.

Brightness Contrast

When a visual stimulus is presented against a background of determined brightness, its brightness contrast may serve as a basis for discrimination. Provided that the background illumination is stable (as it is in the interior compartments of a submarine), and that the eye of the receiver is adapted to that background, contrast discriminations are stable and precise. It is well-established that, within limits, the greater the difference in brightness between a stimulus and its background, the shorter the reaction time to the stimulus when it is first presented. It is probable that at least four values of contrast may be employed in visual communication systems without sacrifice of accuracy or speed of discrimination. As with brightness, contrast is reduced over long ranges.

Color

Color suggests itself as a basis for discrimination among stimuli for the purpose of communication. Provided that the colors employed are selected so that they are not easily confused, either by the color-blind or color-weak, and that they can be discriminated rapidly (**11**), color can be employed as a basis for interior communications.

Very serious difficulties arise in the use of color discrimination for visual communications where any great distance may intervene between the receiver and the stimulus he is responding to.

There are decided differences in the abilities of humans to differentiate among colors. These differences are exaggerated when the discrimination is made more difficult, e.g., when a colored light must be discriminated at small subtense (as in blinker signalling), or where it must be discriminated through a column of air, which has the inevitable effect of reducing the brightness and contrast of the source. Lights of easily discriminable hue, when viewed through a great space of atmosphere, are not accurately and quickly discriminable by a person with normal vision, to say nothing of the color-blind and color-weak.

Temporal Patterning

Temporal patterning of the visual stimulus is an excellent basis for visual communication. It is, indeed, the technique employed in all blinker communications. However, a second type of temporal patterning also suggests itself: the simple flickering, at three or four rates, of the stimulus. Several flash-rates selected, as with contrast and hue, for accuracy and speed of discrimination could readily serve as the signs for an equivalent number of significates. Such flashing lights have the property, not shared by radio-coded temporal patterns, of continuing unchanging until the message is acknowledged, and the appropriate action begun.

Temporal patterning in blinker communication is a special problem. It will be suggested later that it may be possible to increase the rates at which such patterns can be received.

Although the duration of a single flash could serve as the basis of a discrimination, it should perhaps be ruled out of communications, since a single flash might be missed, by blinking of the receiver, or by his faulty orientation. If the flash is repeated regularly in order to avoid this problem, duration reduces to an equivalent of flash rate.

Size

Size of the stimulus is another basis for discrimination. Since the size of the retinal image is a function of both the size of the stimulus and the distance of the receiver from it, size does not recommend itself for communication over variable distances, but can be relied upon if the receiver remains at a relatively constant position with respect to the stimulating display. Thus, size may be employed in interior communication with assurance.

Shape

The shape of a stimulating patch of light—whether it is a circle, square, triangle, or a particular letter—is a convenient and obvious variable property of a stimulus for use in communication. The limiting values of visual acuity reduce the usefulness of pattern for exterior communication, although it is employed in the signal flag.

Position

The stimulating light may appear in any of several designated positions. This variable is often inconvenient for exterior communication.

Movement

The single stimulus light may be made to move back and forth in easily discriminable directions, either actually or by use of the phi phenomenon. Again, this variable does not lend itself to exterior communication.

So far, we have dealt with eight variables, seven of which can easily be employed as a means of communication. If only four values of each of these variables are taken, we have 4^7 different combinations which define modifications of a single spot, each of which may serve as a symbol in interior communications. There is, then, ample opportunity, using only one spatially-continuous visual stimulus, to communicate a large variety of information if the receiver is at a fixed distance from it, and views it against a constant illuminated background. Such a system could be most flexible. Since the number of different messages need never be nearly so great (it will be determined by the traffic and content analysis of the communication system), the number of such symbols actually used would be a small fraction of the theoretical total, and the amount of training required to learn to send and to respond appropriately to such symbols would be within practical limits.

Numerosity

The number of stimuli may be varied. The studies of Reese (**12**) suggest that the upper limit in the number of stimuli for communication purposes should be six, since this is the approximate maximum that is discriminated correctly when presented simultaneously ("subitized").

Spatial Patterning

Spatial patterning of several simultaneous stimuli includes all cases of the use of visually presented words and sentences, and also of dials, pointers, and similar visual displays. The variables which determine legibility and optimal instrument design are pertinent to the use of such spatial patterning in visual communication. A series of dials, indicators, or words presented to the receiver may conveniently serve as the means for establishing man to man couples, just as they do for machine to man couples.

There has been occasion in the preceding paragraphs to note several limitations on the visual dimensions which may be employed in visual communication. The limitations stem from the following physiological functions: (1) the photopic and scotopic visibility curves, (2) the visual acuity function, (3) the level of adaptation of the receiver, (4) the intensity discrimination functions, (5) the hue-discrimination function, (6) the flicker function, (7) the sensitivity functions of the visual field. For exterior communications, where the distance between stimuli and receiver may vary, and where atmospheric attenuation adds yet another variable, it is clear that the limits imposed by (1), (2), (3), (4), and (5) will vary in an irregular way, so that any discriminations set up along these dimensions will yield unreliable communication. Temporal patterning is the only relatively stable method of communication over a great and variable range.

For these reasons, studies of visual signalling have been largely concerned with establishing the ranges at which stimuli of varying brightness and color can be just seen, or just identified with respect to hue, as a function of various meteorological variables.

Where the range of observation is fixed, however, and where atmospheric attenuation is negligible, as in the interior of a vessel, these physiological limitations on the precision of communication do not offer serious obstacles, and the problems of establishing a rapid and accurate system of visual communication become those of designing the system and of training the senders and receivers.

The Present Status of Visual Systems of Transient Communication

Direct Systems

The variety of direct visual communication systems is relatively limited; the situations in which they have been most useful are those in which attenuation by distance has made voice communication impossible, or unreliable without the intervention of a mediating system. Consequently, they have been used in military situations where the sender is at some distance from the receiver, so that the receiver is typically stimulated by the sender's signal through a long column of atmosphere. The optical transmission of this column of air sets an unavoidable limit[3] on the ranges over which direct visual communication can be established and lends it all of the predictable unpredictability of a weather forecast, and all of the ultimate uncontrollability of the weather itself. Consequently, direct systems are becoming progressively less important as radio techniques develop. On the other hand, the utility

[3] The attenuation of electromagnetic radiation by passage through water clearly limits the range over which under-water visual communication can be effected. So sharp is this limit that it is unfortunately unnecessary to devote space to the possibility of submarine-to-submarine visual communication when they are submerged. Experimental work in this direction has proven abortive.

of very few of the various methods has been explored fully in those situations where they may be employed.

The first set of direct systems includes those in which the receiver sees some behavior of the sender that has become symbolic, through the previous training of both sender and receiver, and accordingly is responded to specifically and appropriately by the receiver.

Gesture

Gestures which have been employed as stimuli in communication range from the simple raised hand of the schoolboy who wants to leave the room, to the elaborate series of symbols employed in the refereeing of a game. A few experimental studies have been made on gesture, but these have been restricted in scope. In military situations, gestures have but rarely been employed, owing to the critical lack of data on the fineness of differentiation made by the sender in producing gestures, and of the receiver in discriminating them. Other methods of visual communication have been preferred for their greater precision, although it is undoubtedly true that gesture is capable of greater precision than is now shown.

Sign Language

Sign languages are essentially refinements of gesture from gross bodily movements to movement of specific parts of the body, usually the arms and hands. The sign systems of the deaf indicate the great flexibility of such behavior, as does the development of a sign language like the lingua franca of Indian tribes who did not share a spoken language. Sign language is effective only over very short visual ranges, even where visual aids are available, and is a relatively slow method. The difficulty of learning, as well, necessarily eliminates sign languages, except very simple ones, from serious consideration where auditory communication is impossible.

Lip Reading

Here, the signs of the sender are produced incidentally to voice communication and are effective only over very short ranges, where the receiver, either because of deafness or of high noise levels, is incapable of auditory discrimination. As flexible as language, it is both imprecise and difficult to learn, and consequently is not at all suitable for military exploitation.

The second set of direct systems includes those which we have termed supplemented. Here, the sender directly controls the activities of some device. These activities are directly seen by the receiver. In this class, we find the basic methods of exterior visual communication now employed by the Navy: mirror signalling, searchlight blinker, signal flags and semaphore.

The Heliographic Signalling Mirror

The heliographic signalling mirror was developed to its present status during the recent war as a means of signalling aircraft in air-sea rescue. Its advantages lie in the extreme simplicity of the apparatus, and the ease with which it may be used. However, the information transmitted is of the simplest sort: it is signalling to the receiver of the presence of the sender at the point of origin of the flash produced when the mirror reflects the image of the sun into the receiver's eyes. Although the use of such mirrors has been restricted to air-sea rescue, there is no reason why mirror signalling cannot be employed as an alternative method of blinker signalling when the sun is bright and high. Given the proper solar elevation and good atmospheric conditions, mirror signalling can supplement search-light blinker by day. As a method of communication, then, it can be expected to show many of the characteristics of searchlight blinkers.

Searchlight Blinker Signalling

Searchlight blinker signalling is perhaps the most common of the supplemented visual communication systems. Here, the sender modulates in a temporal pattern the intensity of a light directed at the receiver, to whom the signal appears as a flashing point of light. The pattern of long and short flashes follows the same alphabetical code that is used in auditory code communication (**13, 14**). Like all other direct visual communication, its utility is restricted by the variability of the meteorological visibility range. From the data available, it would appear that all such modulated light beam communication systems have a practical upper limit on speed of communication of some 15 words per minute. The average speed with which messages may be received is approximately 12 words per minute, which is strikingly slower than the speeds obtained with verbal communication, but only slightly slower than that of auditory code.

In blinker signalling a problem arises which is also pertinent to mediated visual communication systems which employ intensity modulation of a stimulating light source: does the present maximal sending and receiving rate represent the highest obtainable maximum? There is reason to believe that this is not the case, but that suitable research procedures might reveal techniques whereby the transmission rate can be increased. First, studies (**3**) of critical flicker frequencies (the rate of flashing which is just too high for the retina to be able to respond to the separate flashes, with the result that the flashing source is seen as a dimmer but steady light) show that the retina, at least, is able to keep up with a much higher rate than that equivalent to the present spacing of dots and dashes in blinker signalling. There are also data which indicate that the critical flicker frequency is not the same for light of different wavelengths, but is higher for the shorter wavelengths, and also that this difference is dependent on the light-dark ratio. It is not unlikely that if suitable investigations were made with the view of relating blinker-reading to critical flicker frequency functions, and the nature of the relationship, if any, were established, our knowledge of the flicker function might be employed to increase both the speed and accuracy of blinker communication.

A second possibility for improving the efficiency of blinker signalling suggests itself in that blinker signalling now makes no use of the position variable in visual stimulation. The use, for example, of four blinking sources, arranged at the corners of a square whose size is determined by the distance of the receiver should make possible the use of simpler codes, requiring less transmission time per letter. The position and number of the points of light, and perhaps illusory movement, in the code, should improve the speed and accuracy of communication without unduly complicating the apparatus.

Blinker has been widely used in recognition signalling. It has proven reasonably effective. However, when it fails, it fails with a vengeance. The flash of the blinker searchlight or Aldis lamp is all too readily mistaken for the fire of light arms at short ranges, and for anti-aircraft fire when seen from great distances. Well-trained pilots, alert to attack, all too often have fallen victims to this easy misinterpretation of a flash, with at least uncomfortable, if not disastrous, consequences for the vessel employing blinker in this way. In Naval warfare, blinker is dangerous as a recognition method.

Signal Flags

Signal flags represent the visual code par excellence. Flags may be selected to symbolize not only letters of the alphabet, but also complete messages. The difficulties encountered are, however, numerous. Not only are flags too slow for most transient communication, but wind often renders control of their mode of presentation impossible. Finally, the demands made by them on hue discrimination render them treacherous at best. Spatial patterning of visual stimuli, such as that suggested above for blinker, is employed in the signal flag, but this advantage is reduced radically by the difficulty, noted above, of making the pattern effective. Signal flags may persist as a colorful, occasionally convenient, and historically interesting system, for certain types of persistent communication, but they have no place in transient communication. They are relics of an earlier age.

Semaphore

Semaphore signalling is essentially supplemented and formalized gesture. Its greatest merit rests in the simplicity of the equipment needed, and, indeed, the dispensibility of any equipment at all in emergencies. However, like all the direct methods of communication, except blinker, it is not effective over any useful range without the use of visual aids by the receiver.

Mediate Systems

The second major class of visual communication systems is the mediate system. Mediate systems vary considerably in complexity of equipment and in flexibility. The simplest of these systems have borne a great similarity to the direct methods. They are dependent upon an unobstructed optical pathway between sender and receiver, and the display is seen through a large column of atmosphere. In the more complex procedures, proximate stimulation of the receiver is employed.

Sea Dyes

The simplest mediate visual system is the use of dyes, usually fluorescent, dissolved in the sea, which serve to signal the receiver of the presence of a sender in the vicinity of the highly colored water which results from release of the dye. Sea dyes were developed for, and have been used most extensively in, air-sea rescue work.[4] Obviously, this type of mediate system is inflexible. The signal is either present, or not, so that no very precise control of a receiver can be exerted through such a simple variation in the stimulus. However, in the sea-dye, we may have a method of visual communication which can be useful for recognition signalling by submarines. In air-sea rescue work, communication at maximum range is desired, so that the color of the dye has been determined by the relative visibilities of dyes of different color. In recognition signalling, however, the signals are seldom if ever given or required at maximal visual range, so that this variable is no longer important. It is suggested, therefore, that the use of fluorescent dyes of different colors, released into the sea in various spatial patterns, be seriously considered as a possible solution to the problem of the recognition of submersibles by air craft. Several advantages over current procedures of recognition signalling are evident. (1) Sea dyes are relatively invisible to surface vessels, although they are conspicuous to aircraft. (2) Areas of dyed water are conspicuous and easily identified by friendly properly briefed pilots. Since recognition signals will be given only when the behavior of an aircraft indicates to the commanding officer of a surfaced or snorkelling submarine that his ship has been seen, and since an unfriendly aircraft will attack, signal or no signal, such signals should achieve their stated purpose, the prevention of attack by friendly aircraft, without risk of the misinterpretation of a flare or light-flash as gunfire, and without revealing the presence of the submarine to all surface craft within visual range. (3) Sea dyes cannot be mistaken for gun fire, as is the case with other methods of visual recognition, such as blinker and flares. (4) By dispersion of dye particles of the proper size, dyeing can be produced rapidly, and may also be effective

[4] The use of colors on life rafts and jackets, and of cloth panels, for the same purpose, is perhaps yet another method of communication.

when a submarine is submerged. (5) Recognition color-patterns can be changed daily.

Release of dyes by a submarine might be brought about in any of several ways. Four small guns, fixed to the snorkel, might fire shells loaded with dyes[5] of the appropriate particle size, in each of four directions. Fuse mechanisms could then explode the shells at a height above the sea surface such that particles of dye stuff are scattered evenly and closely over a wide area. Thus, when a friendly plane approaches, the C.O., firing the four guns by remote control, could produce within a few seconds four large colored patches of water, in a spatial pattern determined by the date and time of day. Such a pattern, viewed from shooting range, offers a minimal basis for misinterpretation by friendly aircraft, despite its use of color. Another method might employ high-pressure jets of heavily dyed water emitted squid-like by the submarine.

Pyrotechnics

Flares and smoke bombs have been employed in visual communication, but because of their lack of flexibility their use has been confined to recognition signalling and to distress signalling. In certain situations they probably will remain indispensable, e.g., flares for emergency use at night and smoke bombs for simple communication from a completely submerged submarine to a surface vessel. Flares have the disadvantage of revealing the presence of a submersible to enemy craft within a wide range. The situations in which pyrotechnics can be used are so limited, and the ease with which they are mistaken for unfriendly activity is so great, that they cannot be considered as a useful communication system except under very restricted conditions.

For exterior communication, all of the methods we have discussed thus far share certain disadvantages. They are dependent upon visibility—the state of the atmosphere. They lack flexibility, being perhaps unduly limited by the codes employed. And they are slow, and involve security risks, in that almost all may be intercepted by anyone within visual range of the sender.

The remaining methods of mediate visual communication present the visual stimuli proximate to the receiver's receptors. In all of these methods, the visual stimulation is presented by some device which has converted into patterns of light some other form of energy, such as an electrical current, electromagnetic radiation outside the visible spectrum, or mechanical energy. These in turn have been set up by the sender, either by his control of radiation as in radar; by his motions "visually" responded to by an electric system, as in television; by his movements, as in the telautograph; or by his control of special electrical circuits, as in the engine-order telegraph. None of these methods are limited by the meteorological visual range, and they are suited to communication between any two stations whether these stations are on different ships, or ship and plane, on ship and shore, or only in different compartments of a single vessel. The various methods will be discussed in the order of their probable utility to the services, rather than in the order of their complexity.

Television

Television systems are ideal for the transmission of visual information of any type over great distances. They enjoy both the advantages and disadvantages of the specific systems whose symbols are transmitted, with added disadvantages peculiar to television itself. These disadvantages are the complexity, size, and unreliability of circuits used. With uniform equipment, and supplemented by auditory communication, television may become a useful method whereby the receiver is stimulated both aurally and visually by the sender, and is placed effectively in direct contact with him. A television system, then, if it could be made

[5] If the dyes developed also had the property of reflecting the radiation employed in radar, they could provide a still further basis for recognition.

practicable, would become an extremely useful method for transient communication.

The Sound Spectrograph

The Bell Telephone Laboratories and other research units have been developing (**1, 9**) over the past few years the sound spectrograph, a device which picks up sound waves, converts them into electrical energy patterns (which may be transmitted over distances by various means, although this has not yet been done), and renders the distribution of energy over the audible spectrum visible as a complex pattern of illuminated areas on a moving phosphor-coated belt. Theoretically, this system renders speech "visible," and, in a sense, it does. However, the "visible speech" which is so ingeniously produced is not easily intelligible speech. Research on visible speech indicates that eventually people may be trained to understand perfectly (i.e., to respond appropriately to) the visual patterns produced by speech, so that auditory communication can be converted with complete faithfulness into visual communication. However, it is already apparent that such training requires not days, but months. Because of the great time required to learn to discriminate the stimuli presented, and because of the bulk (and probable unreliability) of the equipment involved, sound spectrographs are not likely to play any role in military communication in the foreseeable future. Visible speech may prove a useful research tool in phonetics, in speech pathology, and perhaps with further developments, it may become a practical aid to the deaf.

Radio Teletype

This will doubtless prove one of the most effective and rapid methods of handling exterior communications, both persistent and transient. With proper coding methods, it seems to meet the criteria of a good communication system, whether visual or auditory. It does not, on the other hand, lend itself readily to the solution of the problem of interior communication, or to communication with vessels or stations lacking special equipment.

Infra-Red Signalling Systems

In the recent war, methods of blinker signalling employing infra-red radiation were devised (**8**). The receiver was enabled to respond to such blinker by use of a special viewing instrument that converted the modulated infra-red radiation into visible light through the use of electronic optics and a phosphor screen. Such a system is essentially a refinement of ordinary visual blinker, and is subject to the same criticisms as those made in the section on such communication, except those relating to security.

Infra-red reflecting paint, with appropriate viewing devices, has been proposed as a means which might be employed in solving the problem of recognition. However, the difficulties of altering daily the pattern of infra-red reflecting paint on the vessel to be identified, and of persuading an excited pilot to look at his contact through a pair of easily misplaced goggles, make this a not very fruitful suggestion.

Telautographs

The telautograph seems useful in railroad stations, where a bulky mechanical-electric system of reproducing handwriting is under no great stress. In the military situation, the mechanical, spatial, and temporal demands make the reproduction of handwriting at a distance of limited usefulness.

Radar

In the recent war, a surprising amount of visual communication was effected by use of the interference patterns which can be set up on the radar scope. Many vessels, by keying their radar signals according to the conventional blinker and radio codes, were able to set up temporal patterns of interference that could be read like blinker code on the scopes of other vessels employing the same types of radar. Showing many of the limitations of these methods, radar interference communication (in its present state) has the difficulty of being readily intercepted, and of being entirely impossible under conditions of radar silence. This method essentially is a method of "opportunity," useful only on certain relatively unpredictable occasions.

Communication by Instrument Display

The last method of visual communication to be considered may be termed communication by instrument display. This class includes all the various kinds of visual communication employing the visual displays typical of machine to man couples.

At the present time, no use of these is made in exterior communications, but in interior communications a few scattered and poorly designed equipments may be seen. The engine-order telegraph and the torpedo-gyro control panel are examples of such systems. Almost all machine to man couples, however, are of this variety. Dial faces, both clock-face and odometer; simple signal lights, such as the glowing red jewels which indicate closed circuits; the complicated status patterns of illumination of the submarine's "Christmas tree"; all of these communicate the activities and status of mechanical systems to a receiver by the use of symbols.

It is such systems, presenting a panel display of visual symbols, that most effectively utilize the visual capacities of the human for the purpose of communication. None of the restrictions on exterior visual communication that are set by the varying ranges and atmospheric conditions are met here, since the receiver is always within a restricted distance of the panel, and since the interior illumination is close to constant. The visual problems encountered are those already under investigation in studies of machine to man communications, and the results of these may be applied to the special problems of man to man communication.

The limitations which are of importance to such systems stem from the need for training of sending and receiving personnel. After a careful study of the problems involved, it is probable that communication by the display of visual symbols on panels may be effective in improving present communication systems.

Interior Communication by Instrument Display

The problem of communication systems is rather different for the psychologist than it is for the physicist. The problem of the former is that of investigating systematically the various behavioral and sensory means whereby one individual, the sender, may control the behavior of another, the receiver, through the use of symbols. The physicist or engineer is faced with the complementary problem of designing physical systems capable of transmitting the symbols initiated by the sender to the receiver over some distance. The proposals to be made are based on the psychologists' findings, and may be expected to pose special problems to the physicist and electrical engineer. The writer, a psychologist and not an engineer, recognizes that some of the procedures suggested may not be capable of being effected by equipment already designed. However, we shall proceed to make concrete suggestions, with confidence that the electrical engineer can eventually solve the problems that are involved.

Indispensable to the study and development of instrument methods of visual communications is a preliminary study of the communication traffic carried by each of the existing telephone and speaker circuits. A complete classification, census, and content analysis, of all these messages and of their origins and intended destinations, should reveal clear-cut distinctions between necessary[6] and unnecessary communications in content and in direction of messages.

[6] Necessary messages are here defined as those which must be *immediately* and *accurately* received if efficient control of a vessel is to be maintained. They are those in which misunderstanding or delay in transit may have serious consequences for the success of the vessel in its mission.

Analysis of necessary communications should reveal, first, the stations between which messages must be transmitted; second, the minimal content of such messages; and finally, the frequency with which these minimal messages are transmitted. It is the necessary messages which are most frequently repeated with systematic changes only (e.g., "Bridge to maneuvering room: make seven five turns") which lend themselves to visual communication systems.

Such traffic analysis defines the feasible for the designer of the visual system. It yields the total number of different messages which the visual communication system between two particular stations must carry, and so determines the complexity of the visual display by instrument which must be produced at the stations. It also yields the number of symbols that both the sender and the receiver must learn. With the problem so defined, the psychologist is able to decide the specific types of stimuli which must be employed in investigations to determine the display most efficient psychologically, i.e., the display which requires least training of sender and receiver and at the same time yields maximum accuracy and speed of response by the receiver. To the engineers goes the parallel problem of designing reliable multiple-channel circuits for the activation of visual displays.

Specifically, how might visual communication by instrument work, and what advantages might such systems have, as supplements to present voice communications, or as substitutes in the event of breakdown or excessive masking? The systems envisaged would utilize signal lights of varying properties—e.g., dial displays and even, in some instances, perhaps actual words—the exact form taken by the visual displays being dependent upon experimental findings. The stations on such a hypothetical circuit would have identical display boards and message-input equipment, so that each message could be repeated back to the sender. Messages put into the system would also be displayed on all boards until taken off by the receiver to indicate that appropriate action has been taken. In use, the system should have certain clear-cut advantages, which may be enumerated as follows.

1. Such instrument communication can proceed without disturbance by noise. Indeed, it can be disturbed only by electrical breakdowns.

2. The visual symbols employed may be made to persist through a period of time great enough so that the immediate memory span and the span of apprehension do not limit the receiver's accuracy as drastically as they do in auditory communication. The complex messages which may be presented to the receiver by the display panel may remain there sufficiently long for two types of acknowledgment, both of which will ensure that immediate and proper response is made and that the sender is advised of the action taken. The first acknowledgment is a return signal repeating the message back to the sender. The second acknowledgment, executed either when the receiver is beginning to act on the instructions of the message, or when action is complete, entails the wiping clear of the displays of the communication system. This makes possible the instantaneous evaluation of the situation by the sender at all times. Finally, the persistence of the visual stimulus provides sufficient time for the accurate logging of the message, if this is desired.

3. Such visual communication is simple, because it is designed to be so. A preliminary survey of circuits ensures that the system carries only a set number of well-known messages, so that the number of stimuli which must be discriminated is few. This simplicity and designed inflexibility should guarantee that the only communications carried are those for which the circuit is installed and that the problem of circuit discipline is virtually eliminated.

4. A visual display can be observed by several individuals simultaneously. This provides a further restriction on the possibilities of error.

5. Visual systems are silent, except perhaps for bells signalling a message, and should not add to the general noise level where silence is at a premium.

6. Both sending and receiving time are minimal.

These potential advantages should render visual communication systems, which display illuminated signal jewels, instrument dials of a variety of kinds, and other types of visual symbols directly to the receiver, of great utility for interior communications in submarines. They are most fully realized in just those situations in which voice communications become difficult. It is therefore urged that research on the properties and practicability of such systems be intiated, with the view of developing them as methods of communication to supplement many of the voice communication systems now in use.

Conclusion

Improvement in communications may be expected if research is performed on visual techniques. Three areas seem in need of research, and specific recommendations are made with respect to them.

1. The use of fluorescent sea-dyes is suggested as a possible solution of the recognition problem.

2. It is probable that studies of blinker sending and receiving will reveal methods of improving both speed and accuracy.

3. Supplementary visual communication systems should be developed, in addition to the present auditory interior communications, with the view of eliminating many of the difficulties now encountered in voice communication.

References

The list of references given below is by no means an exhaustive survey of the pertinent literature. Those interested in exterior communication by visual signals, on which there is an extensive literature, are referred to Fulton, John F., Hoff, Phebe M., & Perkins, Henrietta T. *A bibliography of visual literature* 1939–1944, and *Supplement*. New Haven: Yale Univ. Press, 1945, pp. 201 + x.

1. Bell Telephone Co. Technical aspects of visible speech. *Bell Telephone System Monograph B-1415*, pp. 89. (Reprinted from *J. acous. Soc. Amer.*, 1946, **17**, 1–89.)
2. Commander submarines atlantic fleet *Standard submarine phraseology*. pp. 24, (RESTRICTED)
3. Crozier, W. J., & Wolf, E. Theory and measurement of visual mechanisms. VI. Wave length and flash duration in flicker. *J. gen. Physiol.*, 1941, **25**, 89–110.
4. Hunter, W. S., & Sigler, M. The span of visual discrimination as a function of time and intensity of stimulation. *J. exp. Psychol.*, 1940, **26**, 160–179.
5. Keller, Fred S. *The radio code research project, final report of Project SC*-88, OSRD 5379, OEMsr-830, Report 8, The Psychological Corp., July 25, 1945.
6. Maninger, R. C. & Knudson, W. T. *Tests on 1MC and 7MC Intercommunication Systems of the USS Perch.* Columbia Univ., Division of War Research, Feb. 17, 1944, (P42/R759). (CONFIDENTIAL)
7. OSRD-NDRC, Division 17. *Transmission and reception of sounds under combat conditions.* Summary Technical Report, Div. 17, NDRC, Vol. 3, Washington, 1946.
8. OSRD-NDRC, Division 16. *Non-image forming infra-red.* Summary Technical Report, Div. 16, NDRC, Vol. 3, Washington, 1946.
9. Potter, R. K., Kapp, G. A., & Green, H. C. *Visible speech.* New York: Van Nostrand, 1947, pp. 441 + xvi.
10. Psychoacoustic laboratory, Harvard univ. *Vocabularies for military communications in noise.* Aug. 25, 1943. (OSRD, No. 1919)
11. Reed, J. D. A note on reaction time as a test of color discrimination. *J. exp. Psychol.*, 1947, **39**, 118–121.
12. Reese, T. W., & Volkmann, J. *The discrimination of visual number.* (Abstr.) Min. and Proc. 23rd Mtg. Armed-Forces NRC Vision Committee. March 4–5, 1949, pp. 85–88. (RESTRICTED)
13. U. S. War Dept., Combined Commissions Board. *Combined visual signaling (V/S) procedure.* FM-24-22, (short title CCBP-5), 12 May 1943. (RESTRICTED)
14. U. S. War Dept. *Combined visual signaling procedure.* (Abridged version for radio telephone (W/T)) FM-24-23, 12 May 1943. (RESTRICTED)

PART V.

HABITABILITY: A. PHYSIOLOGICAL FACTORS

Chapter 13

TEMPERATURE AND HUMIDITY IN RELATION TO THE THERMAL INTERCHANGE BETWEEN THE HUMAN BODY AND THE ENVIRONMENT

L. P. HERRINGTON

Yale University and the John B. Pierce Foundation

AND

J. D. HARDY

Cornell University Medical College

A Foreword on the Habitability of Submarines

A Review of the Physiological Problems

Eugene F. DuBois

Professor of Physiology, Cornell University Medical College, Capt. (MC) USNR (Ret)

The physiological problems connected with submarines are numerous and the publications concerning them extensive. In order to make these publications available the Bureau of Medicine and Surgery of the Navy Department issued in February 1948 a comprehensive bibliography and review compiled by Cmdr. Ebbe C. Hoff[1] (MC) USNR, who not only assembled the references but added a review of each important subject.

Commander Hoff's book covers the field so well that it is only necessary here to mention most of the problems briefly. There is one field, however, that could not be treated in detail because the practical developments were recent and not available in the literature. This is the field of the production of heat in the human body and the loss of heat from its surface. Probably the chief reason that it was not included was the knowledge that a comprehensive volume on this subject was being assembled by the National Research Council Committee on Clothing. Inasmuch as the publication of this book has been delayed Dr. Herrington and Dr. Hardy were requested to prepare the subject in such a form that the principles and formulas would be available to engineers concerned with submarine construction and ventilation.

The most important factors in submarine habitability are heat and humidity. The production of heat in the human body, its metabolism, involves the consumption of oxygen and the production of carbon dioxide. The loss of heat from the body depends upon the temperature and humidity of the environment.

Under ordinary conditions of short submergences in temperate climates these factors are hardly appreciated. It takes a long submergence before the crew has increased the CO_2 or diminished the O_2 in the air to such an extent that the effects are apparent. Arctic cruises point out the need for added heating or clothing and the pre-

[1] Hoff, E. C. *A bibliographical sourcebook of compressed air, diving and submarine medicine.* Bur. Med. Surg., Navy Dept., Feb.1948.

vention of condensation. The tropics show clearly the necessity for air conditioning and the reduction of humidity. This was demonstrated in the Pacific area in World War II.

The main reason can be stated very simply. The human body is maintained at a temperature of 37°C (98.6°F) and the skin temperature is usually 2–5°C lower. When the air temperature is as high as the skin temperature the body is no longer able to lose heat by radiation and convection and must depend entirely on the evaporation of sweat. If the humidity also is high there is no longer efficient evaporation and the heat produced in the body must be stored in the body. This results in a rise of body temperature, first to a point of discomfort, then inefficiency, and finally death.

Beneficial effects of air conditioning were appreciated not only by the improvement of living conditions but also in the avoidance of condensation of water on electrical equipment. The penalties of air conditioning are of course the weight, the bulk and the noise. Efforts must be made to reduce all three of these factors and there seem to be great possibilities. Improvements can be made in the apparatus itself. Some help may be obtained by encouraging the condensation of water on the inner surface of the hull. In the old submarines before World War I there was a great deal of such condensation but it was later eliminated by the introduction of cork lagging. Even in the tropics the sea water is several degrees cooler than the human skin where the sweat is vaporized. If the layer of cork lagging were separated from parts of the hull by an inch, and air were circulated through this space, the air would come in contact with metal almost as cool as the sea water. When the temperature of the skin is 37°C (98.6°F), the saturated air in its immediate neighborhood contains 43.5 grams of water per cubic meter. If the inside of the hull were 30°C (86°F) the saturated air at its surface would contain 30.0 grams of water. This differential would be large enough to reduce the humidity without the penalty of weight or noise. The reduction would not be great but it would relieve some of the load on the air conditioning apparatus and would be of considerable service when silence was imperative.

The chief advantage of separating the lagging from the hull would be a continual movement of cold air downward over the inside of the hull even without the use of blowers. This seems to be the only noiseless method available. There are strenuous objections to this idea on the part of submarine designers and operating officers since it would encroach on space and would lead to rust. The method of running ventilating pipes outboard under the superstructure from one compartment to another would cause great condensation of moisture but would require the use of a noise-producing blower.

The attainment of even partial dehumidification by silent means will demand much experimentation. It is well worth the time and trouble if we can judge from the reports of officers who commanded submarines in the war zone of the tropics. They tell of long submergences with sweat dripping so fast that it filled their shoes and made puddles on the deck.

It must be remembered that the human skin and lungs are not the only sources of water vapor in a closed submarine. Vapor from batteries and cooking contribute to the load, and there may be methods of reducing both of these without much penalty of weight.

It is impossible to emphasize too strongly the fact that it is the vaporization of water from the skin that is of vital importance in regulating body temperature in hot environments. The dripping of sweat from the body is of absolutely no help in losing heat. It merely increases the loss of precious water and salt from the body and eventually leads to the serious picture of water and salt deprivation. Drinking cold water helps relatively little. For example, if a man with a high temperature could possibly swallow a liter of ice water, it would abstract 40 calories, enough to lower his temperature only $\frac{2}{3}$°C. This same water vaporized from his skin would abstract 585 calories.

The humidity of the air can be measured accurately by wet and dry bulb thermometers or psychrometers. Rough estimates can be made by the amount of discomfort and the dripping of sweat. The danger point could be determined by means of clinical thermometers. If several members of the crew are found to have temperatures over 103°F, the situation is critical. In an emergency some relief could be secured by sponging the body with alcohol.

Gases in Submarines

Table A is taken from an article[2] on submarine ventilation published in 1928. In this article it was estimated that under ordinary conditions each man in the submarine produces 0.75 cubic feet CO_2 and consumes 0.9 cubic foot O_2 per hour. Knowing the total air capacity of the submarine and the size of the crew it is easy to calculate the number of hours before the condition will became dangerous. On a long submergence it is desirable to keep the CO_2 below 2 percent and the O_2 above 17 percent until shortly before the end of the dive.

TABLE A

Important Gases Occurring in Submarines

Gas	Formula	Density Air = 1	Dangerous concentrations (approximate)
Carbon dioxide	CO_2	1.53	More than 5 percent.
Oxygen	O_2	1.11	Less than 14–7 percent. Pure O_2 at pressure of 3–4 atmospheres may cause trouble.*
Nitrogen	N_2	.97	Inert; dangerous only after decompression from high pressures in diving suit, etc.
Hydrogen	H_2	.07	Not poisonous; a mixture of about 4 percent may be inflammable; mixtures 8–66 percent explosive.*
Water vapor	H_2O	.62	High humidity depressing; 100 percent humidity, with air temperature above 100°F. dangerous.
Chlorine	Cl_2	2.49	1/100000 to 1/10000.*
Arsine	AsH_3	2.70	1/20000000 to 1/1000000 on prolonged exposure.*
Stibine	SbH_3	—	Supposed to be the same as arsine.
Gasoline	—	—	Mixtures of gasoline vapor 1.4–6 percent in air are explosive; concentrated fumes cause intoxication.
Carbon monoxide	CO	.97	1/10000 in several hours; 1/100 in a few breaths.*

* The table 38-215 in Bureau of Ships Manual gives the following figures for dangerous concentrations:
O_2—Less than 12-7 percent at 29.92 in. Hg. Pure O_2 at pressures of 3–4 atmospheres may cause trouble if breathed for an extended period of time (C. K. Drinker, verbal communication says one hour).
H_2—mixtures 8-74 percent in air explosive or 8-94 percent in O_2.
Cl_2—0.0004 percent for $\frac{1}{2}$ hour, 0.004 percent for a few minutes.
AsH_3—0.003 percent, slight symptoms after several hours exposure, 0.005 percent dangerous for one hour exposure, 0.025 percent fatal in 20 min.
CO—0.01 percent if breathed for several hours, 0.15 percent if breathed for half an hour.
C.K. Drinker, Carbon Monoxide Asphyxia, Oxford Univ. Press 1938 gives the following figures for CO by volume percent:

Gasoline automobile exhaust gas	7.0
Diesel engine exhaust	0.1–0.33

[2] DuBois, E. F. Physiology of respiration in relationship to the problems of naval medicine. Part III. Submarine ventilation. *U. S. nav. med. Bull.*, 1928, **26**, 8.

There is little danger from 3 percent CO_2 and 12–14 percent O_2 for a brief period before ventilating the boat. The 1928 article discusses the physiological principles which are concerned with the most encomical use of materials to absorb CO_2 and replace the O_2. Economy is important in a long cruise and in case of a two or three day submergence can add another day of life while waiting rescue.

In Dr. Hoff's *Bibliographical sourcebook* the subject of air purification is brought up to date. The older method of removing CO_2 by soda lime in canisters has been replaced by lighter and more efficient absorbers, but some of these are caustic and not without hazard. Oxygen is still supplied from heavy tanks. More experiments are needed to show the best methods of absorbing CO_2 and providing oxygen. There have been marked improvements in the transport and use of liquid oxygen.

Although rough estimates can be made by calculation as to the time when it is necessary to start removing CO_2 or supplying O_2, these should be supplemented by air analyses. A good many years ago the submarines were provided with a simple method of carbon dioxide analysis contained in a little box. As far as the writer can tell the apparatus was seldom used in the Pacific during World War II. If submergences are to be prolonged air analysis is imperative in order to economize materials and retain the efficiency of the crews. Even though the apparatus is simple it must be inspected and tested frequently. Great accuracy is not needed but the method should, for example, be able to tell whether the CO_2 is 2% or 3% and the oxygen 16% or 17%.

The danger from hydrogen can be controlled by adequate battery ventilation, but there may be emergencies in which the gas approaches a concentration that is explosive. Elaborate electrical detection apparatus is accurate only if it is in perfect order and calibrated against known mixtures.

Arsine and stibine may in rare instances cause trouble if arsenic and antimony are present as contaminants of battery lead. The tests for them are simple. Chlorine is generated if sea water gets into the batteries. Fortunately the alkaline absorbent of CO_2 also removes Cl_2.

Carbon monoxide may cause serious trouble with any new type of fuel or engine. Here again the tests are fairly simple. The effects of poisoning by carbon monoxide are so insidious that a commanding officer would have little warning before complete disaster.

The odors in submarines, particularly after long submergences, cause great annoyance. The chief source of odor is the human body, but the smell of smoking, cooking, garbage, and the venting of sanitary tanks make heavy contributions. Each factor should be studied by itself and at the same time there must be efforts to absorb or counteract the whole spectrum of stinks. Apparently charcoal as an absorbant has been disappointing.

The food on submarines, well discussed in a recent article by Shilling and Duff[3] has been on the whole satisfactory. There is still the problem as to the method of securing adequate ventilation of food, especially meat, in ice boxes in order to have maximum capacity without danger of spoilage.

It is to be hoped that the engineers will be able to make improvements in the lighting and painting of submarines. They could probably secure more space for personnel if they redesigned some of the bulky instruments and controls. Snorkeling will tend to eliminate some of the exposure on the bridge in cold and stormy weather, but there is still the crying need for clothing that will keep an officer warm and dry.

The snorkel, on the other hand, has introduced its own problems. The sudden fluctuations of pressure are hard on the ears and sinuses and the effects of long periods below the surface have not yet been explored thoroughly. Every new type of submarine, engine, fuel, and oxygen supply must be

[3] Shilling, C. W., & Duff, I. F. Analysis of submarine food problems in world war II. *U. S. nav. med. Bull.* 1948, **48**, 683.

studied with care at the earliest possible stage of its development.

Note on Translation of German Monograph on Submarine Medicine

At the end of World War II the Technical Section (Medical) U. S. Fleet, U. S. Naval Forces, Germany, kept in touch with the staff of the German Submarine Medicine Research Institute after they evacuated the laboratory at Carnac, France. A lengthy monograph covering the work of this Institute was prepared and translated. "Any reserve supply of this monograph remaining after the first distribution will be in the possession of the Research Division, Bureau of Medicine and Surgery. Inquiries concerning further distribution should be directed to that office."

Monograph on Submarine Medicine, Edited by Dr. K. E. Schaefer. The following subjects are discussed, some of them at considerable length: The medical service on the submarines, man and environment, trace substances in the air, carbon monoxide, compartment air, absorption of CO_2 in water, minerals in blood, gases in submarines, metabolism, circulation of blood, heat regulation, adaptation, experiments on guinea pigs, CO_2 intoxication, clinical investigations, effects on ears, "Schnorchel" submarines, nervous syndrome, rapid method of O_2 determination, skin temperature recording. There are numerous case reports of nervous disorders.

END OF FOREWORD

HEAT PRODUCTION OF THE HUMAN BODY

The minimum heat production of the human body is attained under conditions of bed rest after fasting for a period of 8 to 10 hours. This level of heat production, known as the basal metabolism, is near 40 kg. cal./m.^{2}/hr. for the average male adult of mature years. Variations in this basal rate are associated in a systematic manner with age and sex, and the expression of the rate in terms of surface area reduces the total heat production of individuals of different size to a comparable basis. DuBois (**14**) has made an extended analysis of the basic physiology of this process.

The total basal heat production of an adult male of average body size is near 70 kg. cal./hr., which represents the basal rate per unit surface given above multiplied by a total surface area of 1.7 m.^{2} If the strict rest and fast conditions to which these basal rates apply are altered to include the sitting posture, light activity such as reading, and a period of fast 2 to 3 hours in length rather than 8 to 10, the metabolic rate per m.^{2} rises to a value near 50 kg. cal./m.^{2}/hr., or 85 kg. cal./hr. for the average man. This level of heat production (near 100 watts per man) has been designated as the "met" by Gagge, Burton, and Bazett (**16**), and has been widely applied in calculating the clothing insulation required by various combinations of human heat production and environmental temperature.

Effect of Activity on Heat Production

Physical activity is the primary factor responsible for elevating total heat production per man above the 100-watt level (85 kg. cal./hr.) noted above as a reasonable estimate for complete relaxation in the sitting posture. Sherman (**36**) gives the following values for average rates of heat production associated with various kinds of physical activity.

In Table I the rates have been given in basic units and also in multiples of the "met" unit and in watts. The "met" unit is arbitrary in nature, but it does possess the advantage of expressing heat production in simple multiples of a standard rest condition familiar to everyone. Expression in watts is not conventional, but has some advantage for readers more familiar with engineering than with physiological units. It permits the heat addition to closed spaces resulting from human occupancy to be compared with inputs from heater units, incandescent lights, and electronic apparatus.

The figures quoted in Table I make no reference to duration of activity. In estimating the heat stress to which the human body may be subject, or the heat which human activity may add to a closed space, it must be clear that increasing rates are associated with decreasing durations in some manner. A very condensed statement of this relationship can be given.

TABLE I
HEAT PRODUCTION ASSOCIATED WITH VARIOUS ACTIVITIES
(from Sherman, **36**)

Occupation	Calories per Man per Hour	Watts per Man per Hour	Mets*
Sleeping	65	76	.76
Sitting at rest	100	116	1.2
Typewriting rapidly	140	162	1.7
Walking at 2.6 miles per hour	200	232	2.4
Walking at 3.75 miles per hour	300	348	3.5
Stone working	400	465	4.7
Swimming	500	581	5.9
Walking up a flight of stairs	1100	1279	12.9

* Multiples of 85.0 kg. cal./hr./1.7 m.^{2} equivalent to met unit of 50 kg. cal./m.^{2}/hr.

Benedict and Cathcart (**5**) observed a professional bicycle rider working to the point of exhaustion on a bicycle ergometer. He was unable to continue after 4 hours and 22 minutes; and during this period he worked at a rate of 111.6 kg. cal. per hour. The metabolism of the subject was not actually observed; but the excess over the basal metabolism was computed from the work performed, assuming an efficiency of 25 percent. During the period of activity, the subject accomplished work at the rate of 112 Calories per hour, giving a computed total metabolism of 526 Calories per hour, approximately Sherman's estimate for swimming. Henderson and Haggard (**25**) recorded the work performed in a crew race in which exhaustion occurred after 22 minutes. Here, the work performed was at the rate of 288 Calories per hour. Assuming a 25 percent efficiency, this indicates a work-metabolism rate of 1152 Calories per hour, plus an estimated basal metabolism of 80, a total of 1230 Calories per hour.

Finally, Nielsen (**31**) actually recorded the metabolism of a subject working on an ergometer under such extreme conditions that the subject was completely exhausted in 22 seconds. (The presence of the number 22 in these three studies is coincidental.) Nielsen's subject showed a metabolic rate during his 22 seconds of 3930 Calories per hour.

TABLE II
APPROXIMATE TIME TO EXHAUSTION AT VARIOUS LEVELS OF WORK

	Period of Exhaustion, hours	Metabolism, kg. cal. per hour
Benedict	4.367	526
Henderson	0.367	1230
Neilsen	0.006	3930

From these three studies we derive the relationships of Table II.

Comparing Benedict and Henderson, an approximate doubling of metabolic rate decreased the time limit to about one-twelfth of its Benedict value. Comparing Henderson and Nielsen, a further tripling of metabolic rate decreased the period of endurance to one-sixtieth of the Henderson figure.

Thus we see that enormous rates of metabolism can be maintained for brief periods of time. Computations on an hourly basis are, however, obviously artificial. Only some figure between 500 and 1000 Calories could actually be maintained for as long as one hour. On a 24-hour basis, it has been estimated that trained participants in a six-day bicycle race might maintain a rate of some 10,000 Calories per day. Under ordinary circumstances, the metabolism of a 24-hour period with average muscular work has been computed by Sherman as given in Table III.

If we substitute a heavier task such as stone working (400 Calories per hour) under the eight-hour work period, the total would be increased to 4660 Calories per 24 hours. The daily extremes for various forms of normal human activity will vary between 2000 and 5000 Calories.

TABLE III

Total Heat Production of a 24-Hour Period Assuming a Routine of Work and Rest

	Total Calories
8 hours sleep at 65 Calories per hour	520
6 hours sitting at rest at 100 Calories per hour	600
2 hours light exercise at 170 Calories per hour	340
8 hours carpentry work at 240 Calories per hour	1920
24-hour Total	3380

Influence of Environmental Temperature upon Metabolism

Within a range of atmospheric conditions, within and slightly above the comfort range, metabolism remains unaffected by environmental temperature. Beyond this range, metabolism rates tend to increase on each side for quite different reasons. The initial stages of this phenomenon are illustrated by an analysis of the following data obtained at the Pierce Foundation.

In one series of experiments by Winslow, Herrington, and Gagge (**42**) 35 tests were made on each of two nude subjects, at varying operative temperatures from 18°C up to 41°C. (The operative temperature represents the combined effect of air and wall temperatures, and will be discussed later.) Air and walls were nearly identical; under these circumstances operative temperature is, of course, the same as the air temperature.

In another study by Gagge, Winslow, and Herrington (**17**), 44 experiments were made on clothed and on nude subjects, with a somewhat similar variation in air temperatures, and, again, with walls at the same temperature as the air. In Table IV we have averaged the observed metabolic rate (per square meter of body surface) as observed in three major ranges of environmental temperature.

Very few experiments were performed with the nude subjects below 20°C. In the case of Subject No. 1 in Study A, the value of 42 under cold conditions may be unreliable, since there were only two experiments in this temperature range and one of them showed a quite abnormally low metabolism. However, this subject was obese and hence well insulated. Subject No. 2 in Study A, and the clothed subjects in Study B, showed

TABLE IV

Effect of Temperature on the Heat Production of Subjects Seated at Rest

Operative Temperature	Mean Metabolism in kg. cal. per sq. meter		
	Under 20°C	20°–24.9°C	Over 24.9 °C
Study A, Mean Subject I, nude	42	44	46
Study A, Mean Subject II, nude	56	49	51
Study B, Mean Clothed Subjects	53	49	49
Study B, Mean Nude Subjects		47	49

a marked and significant rise in metabolic rate under the cold conditions, and all of these latter subjects were of normal body build. On the other hand, the hot conditions in all cases but that of the clothed subjects of Study B show a slight but appreciable increase in metabolism. It appears that, in general, the minimum metabolism is observed at 20°–25°C, and that below 20° and above 25°C there is a tendency to increase metabolic rate. The increase under cold conditions is clearly a useful adaptive reaction. This increase is related to muscular tension and shivering. No significant reductions in metabolism under warm conditions have been reported for men. In comparison with the benefits conferred by variations in clothing, we may conclude that no human adjustment of basal or resting heat production is of major importance as a practical long-term protection against either extremes of temperature.

A quantitative study of the total metabolism of submarine crews engaged in typical routines would yield a reasonable estimate of the average daily heat production. In the absence of such studies, considerations described above lead to the conclusion that this figure is in the range from 3000 to 4000 kg.cal. per day per man.

TABLE V

Differential Thermal Behavior of Heated Mannikin and Human Subjects When Exposed to Temperatures of 25°, 30°, and 35°C

Unclothed Heated Mannikin			Unclothed Human Subjects		
Air Temp.	Skin Temp.	Internal Body Temp.	Air Temp.	Skin Temp.	Internal Body Temp.
35°	39°	42°	35°	35°	37°
30°	34°	37°	30°	34°	37°
25°	29°	32°	25°	29°	37°
	Gradients			Gradients	
	$T_S - T_A$	$T_B - T_A$		$T_S - T_A$	$T_B - T_S$
35°	4°	3°	35°	0°	2°
30°	4°	3°	30°	4°	3°
25°	4°	3°	25°	4°	8°

Human Temperature Regulation

In the preceding section it has been shown that rest and work are systematically associated with heat production. The energy for this process is derived from food combustion within the body. Hence, it is possible to consider that human occupancy of confined spaces is analogous to occupancy of the space by multiple heat engines, consuming fuel, and producing heat and motion. A characteristic difference between the mechanical and the biological process, however, lies in the low temperature at which human combustion proceeds (near 37°C) and the elaborate and sensitive mechanism within the human body which holds its internal operating temperature near to this fixed point, despite large variations in heat production, and wide swings in external temperature.

The reality of this thermal regulation and its efficiency may be illustrated by reference to Fig. 1 and Table V. Fig. 1 shows a copper replica of the human body. When filled with water and supplied with internal heat equivalent to 0.75 met, the gradient of temperature from the internal water mass through the skin to the outside air in a room at 30°C is given in Table V. By regulation of the fluid circulation rate the gradients of heat loss established may be made quite similar to those found in a resting human subject at the comfortable and thermally neutral condition of 30°C. When the temperature of the environment is raised or lowered by 5°C, every step in the temperature gradient of the heated mannikin reflects in an exact manner the temperature change in the environment. If the human body responded to a change of ±5°C in the environment in the same manner, the individual would be prostrated with heat at 35°C and rigid with cold, perhaps unconscious, at the 25°C condition. However, as one may see in Table V, this is not at all the case. The human data of this table indicate that the deep internal temperature is quite constant over this 10°C range of environmental temperature. If we add to this primary thermoregulatory ability of the nude body the additional range conferred by voluntary choice of clothing and regulation of heat production through activity, the range of external temperature over which body temperature may be maintained near 37°C for limited periods extends at least from −40°C to +50°C. There is, however, a substantial physiological cost attached to such survival adjustments, and maximum human efficiency is preserved only in a narrow range from approximately 15° to 28°C for the clothed subject.

The experienced engineer on examining the gradients of Table V will see immediately that there are two classes of effects operating to keep the deep body temperature constant. He will ask: Since a constant quantity of heat is passing from the region at 37°C to the skin, why does the thermal head vary from 8° at a temperature of 25°C to 2° at 35°C? This mechanism for altering the internal gradient must be an important part of human heat regulation.

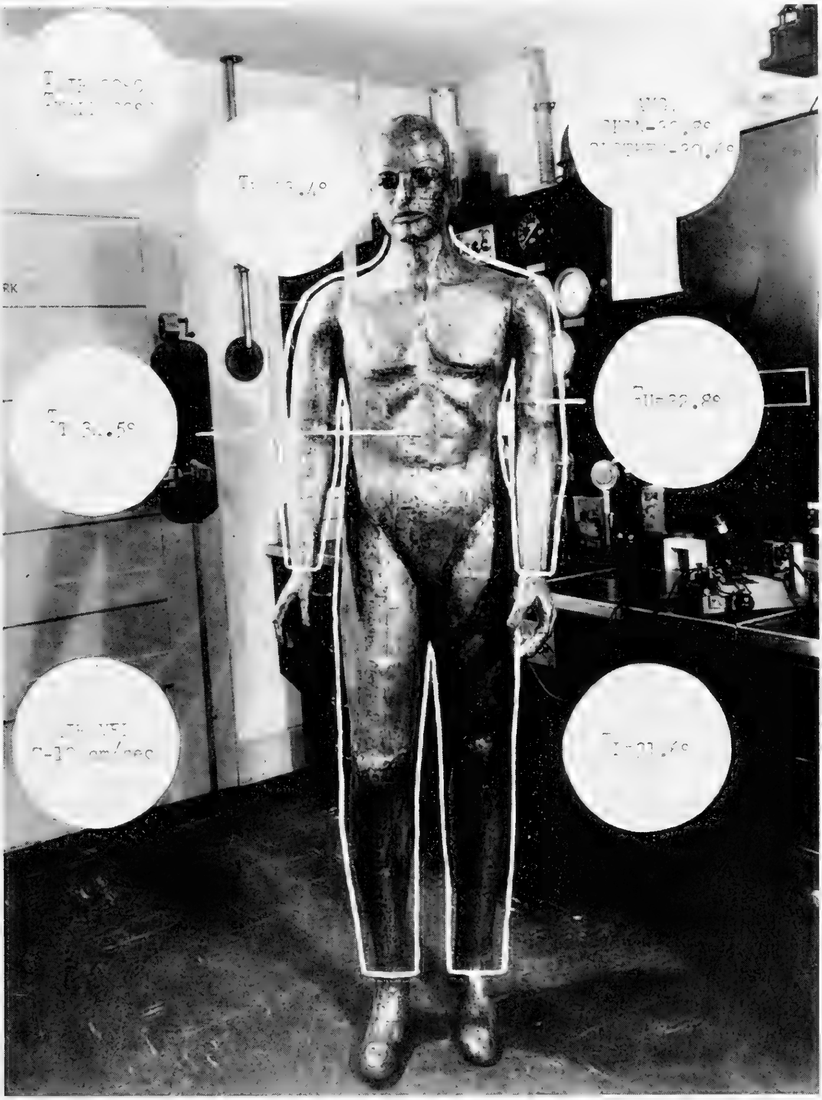

Fig. 1. Electrically heated copper mannikin of 1.8 m.[2] body area whose heat production may be regulated to correspond to either resting or exercising heat production. Devices of this type are frequently used to estimate the protective value of various clothing assemblages. Temperatures shown in °C are skin, clothing, and environmental temperatures, as observed in seated subjects normally clothed and comfortable. Conditions in °F, Air and Wall 72°, T_H (head) 92.1°, T_T (trunk 94.1°, T_U) upper extremities) 91.0°, T_L (lower extremities) 88.9°, Average Skin 91.0°, Average Clothing 87°, Air Velocity 15–25 feet per minute.

This supposition is correct and the

physiologist can supply a sufficient answer. The thermal resistance between the deep body region and the skin can be varied. As in a heat exchanger, the transfer of heat from a circulating medium (blood) to the receiving surface (skin) is a function of the velocity of circulation in the system and the thermal resistance of the heat exchanger wall. By dilation of blood vessels in the skin, the skin thermal resistance may be greatly decreased. Through increased rapidity of circulation in the system generally, and particularly in the dilated skin area, thermal head at constant heat production may be varied by 5° or 6°C. The physiological cost of this adjustment is borne primarily by the heart. The details of such adjustments in circulation index are described by Hardy and DuBois (**23**) and by Winslow, Herrington, and Gagge (**42**).

The engineer will also note that the conditions at the surface of the body as reported in Table V are also different from the mannikin's response. He will then ask: How is it possible for the standard heat production delivered from skin to air over a 4°C gradient at a room temperature of 25° or 30°C to be passed to the air over a 0° gradient at a room temperature of 35°C?

This question is largely rhetorical since the process of evaporation or sweating is generally familiar to all. The physiologist adds that when the body is subjected to a larger stress than can be met by its internal thermal head reduction mechanism, it stabilizes the skin surface at a value 1° to 2°C below internal body temperature by secretion and evaporation of sweat from its external surface. Under these circumstances air temperatures may reach and exceed skin temperature and body temperature.

The Neural Integration of Human Temperature Regulation

We have observed that the body produces heat consistent with different grades of work and eliminates this variable heat input to a variety of thermal environments from a rather constant internal body temperature near 37°C. In a very elementary comparison of the thermal response of a heated mannikin with that of a human subject, we have isolated the gross physical features of this adjustment. It is important to know how these adjustment resources are controlled and integrated.

The primary element in human temperature regulation is a group of cells located by Ranson (**34**) in the hypothalamic area of the brain. This temperature regulating center is directly sensitive to temperature. Due to its location and blood supply, it is constantly perfused by blood whose temperature is a representative sample of the thermal state of the important vital tissues and organs of the body. This center operates as a thermostat set normally at 37°C for resting levels of activity, but capable of resetting itself for higher levels as total heat production increases. Thus, under conditions of strenuous exercise, Nielsen (**31**) has found that 39°C is the approximate control point. In contrast with the usual mechanical thermostats, this biological temperature regulator initiates positive adjustments for excesses as well as deficiencies of temperature, and hence, in a gross sense, resembles the dual control of an air-conditioning system with both heater and cooler units under its control. Not only are the adjustments which this center may initiate to temperature stress influenced by the actual temperature of the cells which comprise the center, but, through its correlative function, it is also influenced by temperature events affecting the numerous warm and cold sensory receptors of the skin surface (Bazett, **3**). Through such inter-connections the center is sensitive both to the slow trend of the temperature of the internal body mass and to the sudden changes which may occur at the body surface. This gives the regulation what would be mechanically regarded as an anticipatory function. Hardy and Oppel have shown that the heat receptors in the skin are sensitive to changes as small as .002°C and the cold receptors to .004°C. There is no direct evidence as to the sensitivity of the heat regulator in the hypothalamus in man, but animal experiments of Ranson and Magoun indicate that this sensitivity is not great; that is, the regulator responds to changes of several degrees in blood temperature.

The adjustments which the center has under control are of two general classes. The first includes the moment-to-moment emergency adjustments which comprise conspicuous features of temperature control. These are dilation or constriction of peripheral blood vessels, alteration of peripheral blood flow, stimulation of sweat secretion, and stimulation of shivering. These actions are not under voluntary control. Actions that require a voluntary element, but are basically regulative, are impulses to alter posture, to increase or decrease food intake, and to alter levels of physical activity or clothing.

In addition to these immediate reactions, the center is involved through complex nervous inter-connections with longer-term acclimatization responses. In this category may fall adjustments in blood volume, in blood chemistry, and in endocrine activity. In animals there are changes in thickness of coat in response to temperature, and, probably in both animals and man, fat deposition is influenced by these longer-term effects.

The Physical Processes of Thermal Interchange with the Environment

In preceding sections the range of human heat production has been indicated and the general effectiveness of temperature regulation has been related to simple processes under the control of a central regulating center. In the thermal control of environments the objective is generally to reduce these adjustments to a minimum. This means that it is necessary to relate quantitative measures of the total heat of the environment to measures of the adjustments which the conditions provoke in the individual. In order to accomplish this a clear understanding is needed of the basic physical processes by which the human body exchanges heat with the environment.

In thermal exchanges between the body and its surroundings, there are four major factors in the picture. The details of this heat interchange as studied by calorimetric methods have been reported by Hardy and DuBois (**18**, **22**, **23**) and Winslow, Herrington, and Gagge (**40**, **42**, **43**). The human body in a state of equilibrium with its thermal environment produces heat by metabolism. It loses heat by evaporation. It loses or gains heat by conduction, convection, and radiation, depending on the environmental conditions. The whole closed system, when equilibrium exists, may be expressed by a simple formula which combines conduction and convection under the latter term.

$$(1)\quad M\ (\text{metabolism}) - E\ (\text{evaporation}) \pm C\ (\text{convection}) \pm R\ (\text{radiation}) = 0.$$

States of disequilibrium or temporarily imperfect adjustment often exist and, in some respects, may be of greater assistance in understanding the problem than the condition of full thermal adaptation. In such states the right side of the equation is not zero, but has a positive or negative value, representing actual chilling or heating of body tissues. This subject will be discussed later.

The nature of the physical processes involved in this interchange should be reviewed, and a condensation of a report by one of the authors (Hardy (**24**)) will be given here.

Heat Transfer by Radiation

By radiation is meant the exchange of thermal energy between objects through a process which depends only upon the temperature and the nature of the surfaces of the radiating objects. Therefore, the flow of heat by radiation does not depend upon the presence of an intervening medium, and heat will pass by the process of radiation from a hot object to a cooler one through a vacuum. Fig. 2 shows diagrammatically the relative positions of some of the various parts of the spectrum, and, as examples, the energy distribution of the radiation from the sun, a red-hot stove, and the human skin, as given by Blum (**6**).

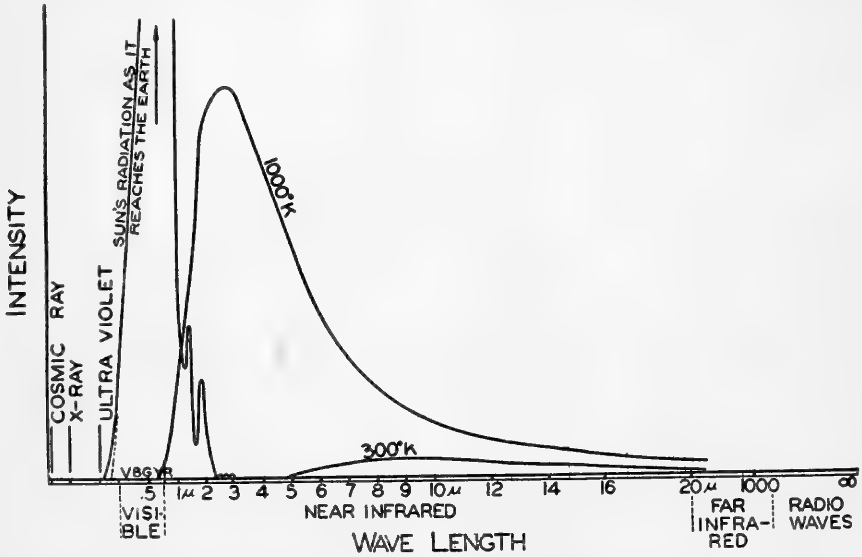

Fig. 2. Relative position in the spectrum of radiation from various sources

Laws of Radiation

The amount of energy which is radiated at wavelength λ, and included in the small spectral range, λ, is given by the law of Planck for black body radiation:

$$(2) \qquad H_\lambda = \frac{C_1 \lambda^{-5}}{e^{C_2/\lambda T} - 1},$$

where $C_1 = 4.93 \times 10^{-15}$ (λ expressed in cm.), $C_2 = 1.432$ cm. degrees, and T = absolute temperature. If λ is large, then the minus unity in the denominator may be neglected and we have:

$$H_\lambda = C_1 \lambda^{-5} e^{-C_2/\lambda T},$$

which is Wien's Law. By simple differentiation of Wien's Law with respect to λ (holding T constant) it can be shown that the wavelength of the maximum radiation is

$$(3) \qquad \lambda \text{ max. } T = \text{constant} = 0.288 \text{ cm. degrees.}$$

This formula is known as "Wien's displacement law" because, as the temperature of the radiating object increases, the wavelength of the maximum radiation is displaced towards the shorter wavelengths. Thus, in Fig. 2, the sun's radiation (T = 6000° K) has its maximum at about 0.5 μ, whereas the hot stove (T = 1000° K) has its maximum at about 3 μ and the human body (T = 300° K) radiates maximally at 9.5 μ. Consideration of Wien's displacement law (together with the eye's visibility curve) shows why one should expect the sun's radiation to appear white, that of the stove red, and that of the human skin to be invisible.

Adding up the values of H_λ for all wavelengths gives the Stefan Boltzman Law for total radiation:

$$(4) \qquad H_B = S_o T^4.$$

The net transfer of heat is the difference in the radiation emitted and the radiation received. Thus we may write for this transfer:

$$H_R = S_o \, \epsilon_1 \, \epsilon_2 \, (T^4 - T_o^4) \, A, \quad (5)$$

in which H_R = heat transfer in gm. cals./sec., S_o = Stefan Boltsman constant = 1.37×10^{-12} gm. cal./sec./cm.^{2}, T and T_o = absolute temperature of the hot object and its environment, ϵ_1 and ϵ_2 = the emissivity of the surfaces of the radiator and environment, with maximum values of unity, and A = the effective radiating area of the hot object.

Emissivity: The Black Body Radiator

By definition, a "black body" is one which reflects none of the radiation which strikes it. That is, it absorbs completely radiation of all types and reflects nothing. The term "black body" is thus quite appropriate. No physical object is completely black in this sense, since every object reflects some light, even though it be a small amount. In contrast to the "black body" is the perfect reflector, a type of surface which is approached in nature by highly polished metallic surfaces (see Table VI). If an object reflects a small but equal amount of light of all wavelengths, it is termed a "gray body." Most surfaces are black, or almost so, for some radiations and not black for others. Such surfaces are colored, and it is to this class of surfaces that the human skin and most clothing belong. The white skin reflects visible light well and does not reflect at all well the invisible infra-red light. Neither negro nor white human skin is by any means "black" for all types of radiation. In fact, the white skin reflects about 30–40 percent of the sun's radiation, whereas the dark skin of the negro reflects less than 18 percent of these rays, according to Oppel and Hardy (**32**). However, in the far infra-red, the region in which the skin reflects little, the skin is a good radiator. Therefore, for all practical purposes, human skin, regardless of its color, can be taken to be approximately black body *for the range of wavelengths in which it radiates.*

Another important factor is the emissivity of the environment. For example, if one were in the center of a large sphere which had highly reflecting walls, nearly all of the radiation emitted from the body would be reflected back to be absorbed by the skin. Therefore, even though the wall of the sphere were very cold, the skin could lose little heat to it by radiation.

For indoor environments in which the walls are at air temperature and are painted, a survey of the walls with a radiometer (**19**) will give their temperature, and ϵ_2 may be assumed to be unity for practical purposes. In more complicated environments, in which hot objects or other radiators are present, resort must be made to instruments such as the thermointegrator (**39**), or the Globe thermometer (**4**). Because of the simplicity and utility of the Globe thermometer as a means of estimating the combined thermal effect of air temperature, wall temperature, and air movement, a reading of this reference (**4**) is particularly recommended for those concerned with air conditioning in undersea craft.

TABLE VI

TABLE OF EMISSIVITIES

(Low Temperature)

(λ max $\sim$ 9–10 μ)

Inside Blackened Cone	1.00
Human Skin	.99
Roughened Rubber	.98
Black Lacquer (glossy)	.95
White Lacquer (glossy)	.95
Copper (polished)	.10 –.15
Silver (polished)	.02

The Radiating Surface Area

It is the body area that is open to the environment that is effective in contributing to heat loss by radiation, since the skin areas between the fingers, under the arms, between the legs, under the chin, etc., radiate to adjacent skin areas and therefore not to the environment. Thus, although the entire surface area of the human body can be calculated from the height-weight formula of DuBois, only a portion of this area can be completely effective in losing heat by radiation. It is obvious that the bodily pose of the individual is of the greatest importance in considering the radiation area, because when a person is curled up with the knees drawn against the chest, the effective area will be much less than in the fully extended "spread eagle" configuration. By means of partitional calorimetry, Winslow (**44**) and his colleagues found a value of 71–75 percent for a man sitting in a chair in a semi-reclining position. Investigators at the Russell Sage Institute of Pathology (**20**) found by two methods a value of 78 percent for a man in a "mummy" position. It can therefore be assumed that for normal extended positions of the body the effective radiating area will be somewhere between 70 and 85 percent of the area figured by the DuBois formula.

Representative values for the emissivities of a number of different materials are given in Table VI.

It is possible to state an equation for computing the heat transfer by radiation. The formula for transfer between nude man and the environment, assuming the emissivity and temperature of the environment can be measured, is as follows:

$$(6)\quad H_R = 1.37 \times 10^{-11}\,(T_S^4 - T_c^4) \times t \times A \times f \times \epsilon, \text{ kg. cal./hr.}$$

where T_S = average skin temperature (°C + 273), T_c = average radiant environmental temperature (+273), t = seconds in one hour, A = DuBois surface area, f = ratio of effective radiating surface to the DuBois surface area (0.78 for nude man lying in anatomical position), and ϵ = emissivity of the environment.

The average skin temperature can be obtained by measuring the skin temperature of various areas over the body surface and weighting the readings according to the proportion of the DuBois (**14**) body area represented. If a man is clothed, the "radiant" surface temperature of both skin and clothing may likewise be determined by weighting temperatures determined by a radiometer. The difficulties associated with surface temperature measurements are critically discussed by Murlin (**29**).

The average radiant temperature of the indoor environment is measured most easily by pointing a calibrated radiometer to many points over the entire solid angle of four radians and making an appropriate average. There is as yet no instrument available which will give this average in a single reading.

The factor f is dependent upon the amount of clothing and the position of the individual; it has been worked out only for the nude man in one or two instances. As mentioned above, the emissivity of the environment is difficult to measure. It is generally safe to assume a value of about 0.95 for most environments, but in cases of doubt, where polished surfaces are present, it is better to make actual temperature measurements of the surfaces in the environment when practical. There is no instrument at present which will give in a single measurement the emissivity of the environment.

In summary, the determination of the radiation exchange between a man and his environment requires six measurements:

a. Physiological

1. Skin or surface temperature
2. Effective surface area
3. Reflecting power of the skin and clothing

b. Environmental

1. Average radiant temperature (infrared)
2. Emissivity of the environment
3. Intensity of radiant sources

Conduction

The flow of heat through a medium without the physical transfer of material is called thermal conduction. Conduction, in contrast to convection, can take place in solids as well as in gases and liquids, although in the latter media appreciable conduction occurs when the circumstances are such as to prevent or limit convection. Heat is conducted from within the human body to the skin surface and from the skin into any cooler objects with which the body may be in contact. The heat lost from the normally clothed body is probably largely conducted through the thin air layers between the skin and clothing.

Laws of Thermal Conduction

In the case of a medium with uniform physical properties, it has been demonstrated that the amount of heat which flows from a warm surface to a cool one is directly proportional to the length of the path, the nature of the medium, and the thermal gradient. Putting this in the form of an equation, we have as the fundamental equation for heat conduction in the steady state:

$$(7) \qquad H_D = \frac{KA(T_2 - T_1)}{d} \times t, \text{ gm. cal.,}$$

where H_D = quantity of heat conducted, K = thermal conductivity, a constant which depends upon the material, A = area of the conducting surfaces, T_2 and T_1 = temperature of the warm and cool surfaces, t = time, and d = thickness of the conductor.

This formula has been applied by many authors to the problem of the conduction of heat from the interior of the body to the skin. Herrington *et al.* (**26**) and Hardy and Soderstrom (**21**) have shown that, when the peripheral blood vessels of normal men are fully constricted, between 9 and 10 kg. cal./m^2/hr./°C are conducted through the body tissues. This insulation corresponds to a layer of tissue 18–22 mm. in thickness.

The equation expressing the conduction heat is entirely analogous to Ohm's Law for electrical circuits, and it has been shown that the conduction of heat through layers of different materials can be expressed in terms similar to those for electricity. In the heat conduction equation,

$$(8) \qquad H_D = \frac{(T_2 - T_1)}{\frac{d}{KA}},$$

$T_2 - T_1$ is the thermal gradient and corresponds to voltage, H_D is the rate of heat flow corresponding to current, and $\frac{d}{KA}$ is the resistance factor, and $\frac{KA}{d}$ is the conductance factor. It is this factor which is important when considering the insulation provided by layers of clothing of different sorts.

A further discussion of conduction as applied to the problem of clothing is contained in a later section concerned with the use of an overall conductivity unit for clothing as introduced by Gagge, Burton, and Bazett (**16**).

To measure heat loss by conduction when a steady state of heat flow has been attained, four measurements are necessary:

a. The skin temperature (T_S) over the area of skin in contaet with the conducting medium,
b. The skin area (A) in contact with the conducting medium,
c. The thickness (d) of the conductor,
d. The temperature of the conducting medium.

The value of the specific thermal conductivity of the medium can probably be found in one of the numerous physical tables.

Convection

Convection is a term which refers to the exchange of heat between hot and cold objects by the physical transfer of the liquid or gas with which the objects are in contact. This type of heat transfer depends upon the actual streaming movement of warm molecules from the warm object to the cooler one. The transport of heat by such streams (convection) is to be distinguished from heat conduction through a gas or liquid, which is the type of heat transfer in which there is no streaming and in which the molecules of the medium remain essentially in the original locations, passing the heat energy from one to another by molecular vibration and collisions. The heating of a building by hot air and the cooling of a surface with an electric fan are examples of *forced* convection, whereas the streams of air rising about a warm surface, the human body for example, are called *natural* convection. Natural and forced convection can contribute to heat transfer at the same time, although it is likely that natural convection will become decreased as the forced convection is increased.

In this discussion we shall confine our description to forced convection, since under the conditions which normally affect human heat loss there is usually some degree of forced air movement. This treatment is also recommended by the fact that in experimental studies on humans it is most feasible to determine by difference methods an empirical convection constant as a function of measured air movement. This constant yields the total convection loss when multiplied by the temperature difference between unit area of a subject and the ambient air.

Forced Convection

An analytical consideration of the basic concepts concerning forced convection will naturally lead to a statement of the factors upon which this heat-loss channel quantitatively depends. It is obvious that the greater the velocity of the air stream, the greater the difference in temperature between the gas and warm surface, etc., the greater will be the convection. Stating this mathematically, we can say that the convective heat, H_c, is a function of these several variables. That is,

$$H_c = f(D, V, \mu, \rho, \Delta T, K, C_p, t), \tag{9}$$

where H_c = heat loss by convection, f = the functional relationship which is not known a priori, but can be determined from experiment, D = the characteristic dimension of the object (for example, the diameter of a sphere or a cylinder), V = velocity of the gas, μ = viscosity—a factor concerned in the mobility of the gas molecule, ρ = density, K = thermal conductivity, C_p = specific heat, T = temperature difference between the warm surface and the air, $\Delta T = T_S - T_A$, and t = time.

How these values can be combined to make a sensible formulation of convection is largely a matter of experiment, but it is also a matter of clever guessing. The final arrangement which best fits both the experimental facts and theoretical considerations is:

$$H_c = \frac{K}{D}\left(1 + a\left(\frac{DV\rho}{\mu}\right)^{\frac{1}{2}} + b\left(\frac{DV\rho}{\mu}\right)\right)\Delta T \cdot t, \tag{10}$$

where a and b are constants depending upon the particular units used. It is convenient to reduce all surfaces to equivalent cylinders or spheres, since most of the experimental work has been done by engineers who are interested in convection losses from pipes. Neglecting everything but convection, the adult human body loses heat like a cylinder 7 cm. (c. 3 inches) in diameter (a = .407, b = .00123, if velocity is expressed in miles per hour, diameter in inches, and convective heat loss in kg. cal./m^2/hr./°F) or like a sphere 15 cm. in diameter (**33**), and the study of convection losses from the formulae developed by the engineers can be most easily made by considering the body as a 3-inch cylinder or a 6-inch sphere.

$\frac{H_D}{K}$ is called Nusselt's Number (NU) by engineers, but it can be seen that, neglecting free convection (and the small last term, $b\left(\frac{DV\rho}{\mu}\right)$), one can write

$$NU = a\left(\frac{DV\rho}{\mu}\right)^{\frac{1}{2}}.$$

$\frac{DV\rho}{\mu}$ is called Reynolds' Number (Re); hence we can say that, for forced convection,

$$NU = a\,Re^{1/2}. \tag{11}$$

The usual method is now to plot Nusselt's Number against Reynolds' Number and thus obtain the value of a. Büttner (**10**) has done this for spheres of several diameters, and finds

$$a = 0.70, \text{ or}$$

$NU = 0.70\ Re^{.52}$, if V is greater than .2 m./sec., or

$$H_c = \frac{0.70}{D}\left(\frac{V.D\rho}{\mu}\right)^{.52} \text{cal./cm.}^2\text{/min./°C.}$$

Doing the same thing for men lying on their backs on the floor, he found that

$$H_c = 0.021(V)^{1/2} \text{ cal./cm.}^2\text{/min./°C.}$$

for the human body in general regardless of size.

Winslow, Herrington, and Gagge (**41**) have determined the convection constants in much the same way and have arrived at the following formula for convection loss:

$H_c = (6.51 + 0.17)\ \Delta T$ kg. cal./hr., for V measured in ft./min., or

$H_c = 2.3\ \sqrt{V}\ \Delta T.$

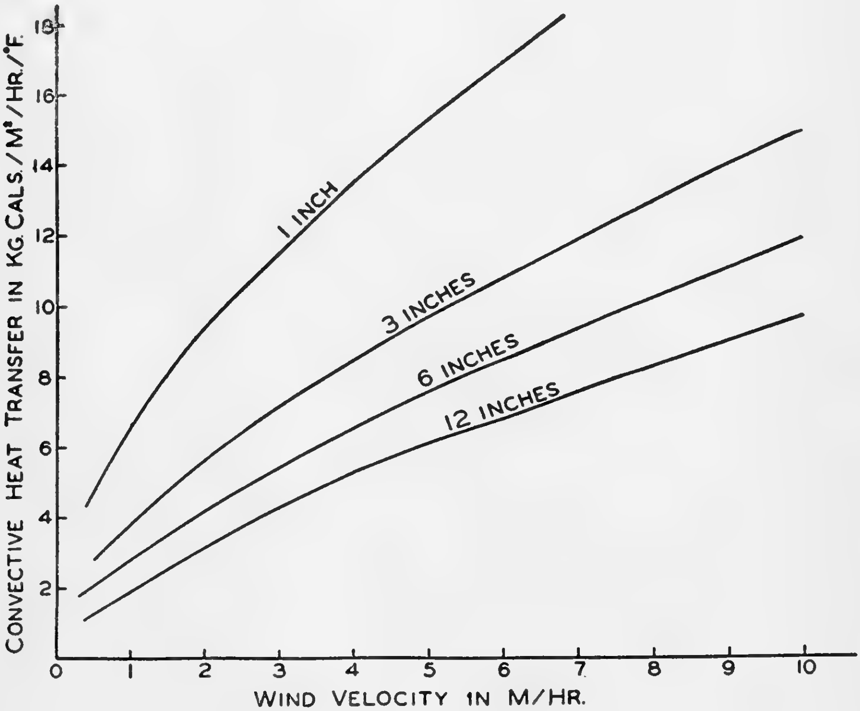

Fig. 3. The convection loss of a cylinder at various wind velocities

Either of these formulae appeared to fit their data. Although the air velocities used by Büttner (**10**) were higher than those used by Winslow *et al.* (**41**), the values of convection are in good agreement.

The formulae given above are valid only for a limited range of air velocities. Büttner (**10**) extended his measurements to 10 m. per sec., and it is at about this velocity that he found the relationship between Nusselt's and Reynolds' numbers broke down. Hence, this velocity may be tentatively assumed to be the upper limit of application for these equations.

During the last few years Plummer (**33**) has made an extension of Winslow's data (**41**) by reducing the characteristic dimensions of the human body to a cylinder 7 cm. in diameter. The curve given below is his plot of the convection loss of a cylinder for various wind velocities.

Two physiological measurements and two physical measurements are required to compute convection losses. The physiological measurements are: (a) the skin or surface temperature, and (b) the characteristic dimension of the body. The physical measurements, which are relatively simple to make are: (c) the air temperature, and (d) the velocity. The measurements of skin temperature will be difficult to make, since the thermometers or other measuring elements must not affect the skin's being cooled by the moving air, and they must not themselves be affected by the air stream. The measurement of the characteristic dimension will require a series of measurements of convection loss by difference as affected by air velocity.

Evaporation

The continual loss of weight from the body has been of interest to physiologists for about 300 years, and the mass of experimental data that has been gathered by many workers under varied conditions has established the importance of this loss as a factor in the heat regulation of the body. Soderstrom and DuBois (**37**) have determined that about 25 percent of the metabolic heat is carried away from the body by the water evaporated from the skin and lungs. The importance of the physiological control of the water evaporated from the skin has been repeatedly pointed out by Rubner, Hill, Kuno, Newburgh, and many others, and in the last twenty years the physical laws concerned with the loss of heat by evaporation from the body have been carefully worked out by Büttner (**9**), Gagge (**15**), and others. Evaporation takes place at the skin surface, the liquid sweat passing into vapor at about skin temperature and finally passing into the environment to cool and to expand to the saturation of the environment.

Although moisture is lost from both the skin and lungs, from the standpoint of heat loss the former is the more important. Winslow *et al.* (**43**) measured the heat loss from the lungs of their subjects and obtained a value of 7–8 kg. cal./hr., about 10 percent of the total heat loss in the neutral temperature zone. Sweating from the skin is the important means of losing heat at elevated environmental temperatures.

Büttner (**9**) pointed out the importance of the amount of skin which can be considered wet, and of the vapor pressure difference between saturated vapor at skin temperature and the vapor pressure of the environment. Gagge (**15**) has given a formula for heat loss by vaporization which is similar to an earlier one by Büttner (**9**):

$$(12) \quad H_V = (\omega\mu)A(E_s - RHE_A) \text{ kg.cal./hr.,}$$

or

$$(12a) \quad (\omega\mu) = \frac{H_V}{A(E_s - RHE_A)},$$

where ω = fraction of body area that is completely wet, μ = proportionality factor containing the vaporization constant and the factors which depend on air velocity and direction, H_V = heat loss by vaporization, A = total body area, and RH = relative humidity.

It is seen from Gagge's equation that when the vapor pressure ($RH \times E_A$) of the air equals that of the skin, no vaporization will take place, and if the air vapor pressure should be greater than that on the skin, moisture will condense on the skin and the body will gain heat.

Hardy and DuBois (**22**), and Winslow *et al.* (**43**) have shown that within the zone of evaporative regulation, relative humidity does not change H_V. This is understandable, as Gagge points out, since the change in the vapor pressure factor $(E_s - RHE_A)$ is compensated for by an increase in the wet area of the skin.

The effect of wind upon the wetted area and upon H_V will, of course, depend more upon the change in total heat loss than upon the immediate alteration in H_V caused by suddenly blowing air upon the body. That is, if the skin is sweating with a value of $\omega\mu$ of 50 percent in quiet air, turning on an electric fan will cause a large initial fall in $\omega\mu$ with a subsequent readjustment to a new $\omega\mu$ which will just let H_V balance the heat equation. It is for this reason that the laws of vaporization relating air velocity, temperature, vapor pressure, etc., as determined from physical analogues of the skin, are valid only under conditions which are predetermined by observations made on the skin.

Formulae 12 and 12a are deceptive in their simplicity, and the physiological variable of the wetted surface should not be lost sight of, since the only practical way to measure $\omega\mu$ is to measure H_V. Fortunately, H_V can be measured rather easily by observing the weight loss of the man, and it is necessary to do this under the various conditions of environment before $\omega\mu$ can be determined.

As seen in Fig. 4a, evaporation in the physiological zone of temperature regulation is not affected by humidity because $\omega\mu$ compensates to keep H_V independent of humidity (Fig. 4b). Gagge showed that $\omega\mu$ has a maximum value of about 28.5 kg.cal./m.²/hr./cm.Hg. for two nude subjects and that this represents a 100 percent wetted skin. By measuring $\omega\mu$ for other conditions of the atmosphere, the degree of wetness of the skin can be calculated by

$$W\,\% = \frac{\omega\mu}{28.5} \times 100.$$

Winslow, Herrington, and Gagge (**43**) have done this for the nude man for high and low humidities and for environmental temperatures between 16°C and 42°C. The evaporative loss is approximately the same throughout their experimental range, regardless of the relative humidity.

In summary, one may conclude that it is not possible at present to measure the ther-

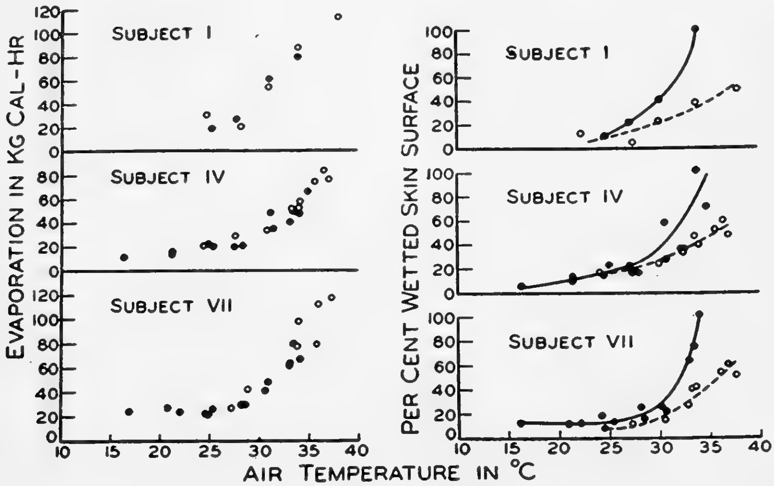

Fig. 4. 4-a (left), Evaporative heat loss in clothed subjects in relation to air temperature. 4-b, (right), Percentage of maximum possible area of wetted skin surface (w) in relation to air temperature (solid circles high humidity, open circles low humidity).

mal factors of an environment by means of physical instruments alone and thereby predict in every case the value of the evaporative heat loss of a man. This is due to the fact that the wetted area of the skin is a variable physiological factor, the behavior of which is not known for all environmental conditions. For the conditions in which the wetted area is known, the evaporative heat loss can be predicted from the measurement of air temperature and velocity, relative humidity, skin temperature, and the characteristic dimension of the body for convection. The cooling effected by evaporation may be computed from the weight of water evaporated and the latent heat equivalent of water at the temperature of evaporation.

The Adaptations of the Human Body to Varying Thermal Conditions

Our discussion of the characteristics of human heat production and of the physical nature of the processes involved in heat interchange with the environment has prepared the way for a description and analysis of the thermal behavior of man when exposed to various ambient conditions. In this section material dealing with the unclothed body is stressed. Similar material for the clothed body is available in the reports of Winslow, Herrington, and Gagge (**44**), and Gagge, Winslow, and Herrington (**17**). However, since in undersea operations heat stress is less well understood and perhaps more critical than cold stress, the reactions of subjects wearing only underwear, lowers, or supporters are of greater value than data on subjects fully clothed in civilian dress.

Combined Influence of Various Environmental Factors

Numerous attempts have been made in the past to simplify the measurement of the various environmental factors influencing human comfort. In England, the Dufton Eupatheoscope (**12**) and the Vernon Globe Thermometer (**4**), and in this country an instrument called the Thermo-Integrator (**39**) were devised in the hope of recording a single figure which would indicate the combined influence of air temperature, air movement, relative humidity, and mean radiant temperature upon human comfort. All these attempts have been abandoned as the various investigators discovered that the physiological reactions of the human body vary so widely at different points in the temperature scale that no single physical instrument can provide a true picture of the environmental influences concerned. Only an independent determination of the four distinct factors mentioned above can give a real measure of the thermal demands of the environment (**35**).

The Research Laboratory of the American Society of Heating and Ventilating Engineers at Pittsburgh approached the subject from another standpoint. Its investigators measured the temperature, velocity, and relative humidity of the air (ignoring radiation effects, since air and walls were at the same temperature). They then obtained from their subjects votes as to the relative comfort of the conditions maintained, and prepared a graph of "Effective Temperatures" representing (at several rates of air velocity) the combination of air temperature and relative humidity yielding equal sensations of comfort. These graphs have become standard for practice in air conditioning, and have given reasonably satisfactory results under hot conditions. There is serious question, however, as to their validity under environmental conditions which are in the cool to moderate range, and recent modifications have been suggested which should remedy the previous undue allowance for the warming effect of high humidities in the moderate comfort range.

At the Pierce Foundation a figure representing the combined influence of air temperature and mean radiant temperature has been found of value, since the physiological reactions to these two factors are so similar (as is not the case where evaporative response to relative humidity is in question).

They derived a factor referred to as "Operative Temperature," defined as the sum of the radiation constant multiplied by the mean wall temperature, and of the convection constant multiplied by the mean air temperature, divided by the sum of the two respective constants. The operative temperature is not a physical condition, since it is weighted by the factors of convection area, radiation area, and air movement; but a measure of heat demand which allows for the physiological reactions involved. When air movement is moderate (25–30 feet per minute), however, the convection and radiation constants are very close to each other, and operative temperature is not radically different from the mean of air and mean radiant temperatures. For the conditions which obtain in ordinary indoor spaces, the engineer can safely assume a mean between air and wall temperatures as representing operative temperature. When operative temperature occurs in the discussion, the reader may assume that the condition is subjectively similar to the thermal condition he associates with an air temperature of the same value.

The General Physical Phenomena of Heat Interchange

The thermal interchanges between the body and its environment are expressed by the formula:

$$(13)\quad M - E \pm C \pm R = \Delta H, \text{ or}$$
$$M - E \pm A_R K(\Delta T_W) \pm A_C(\Delta T_A) \sqrt{V} = \Delta H.$$

All units are expressed in kg.cal. per hour. M = observed rate of metabolism, E = rate of cooling due to the sweat actually evaporated, A_R = the effective radiation area for a given subject in a given position, in square meters, $K = (4kT_S^3)$ from the first approximation of Stefan's law (k being the universal radiation constant), ΔT_W = difference between skin temperature and mean radiant temperature in 0°C, A_C = the convection constant for a given subject in a given position, ΔT_A = the difference between skin temperature and air temperature in 0°C, V = mean turbulent velocity of air movement in cm./sec., and ΔH = change in heat content due to shifts in mean body temperature.

The actual relative magnitude of the heat interchange by various avenues may vary within wide limits. It is sometimes stated that radiation accounts for about two-fifths of the heat loss from the body, convection for two-fifths, and evaporation for one-fifth. This is approximately true for a resting subject at low environmental temperature with little air movement, low relative humidity, and air and walls of about the same temperature. The mean of six of our experiments with a nude subject with both air and walls at 24.3°C showed that 21 percent of the actual heat lost to the environment was due to evaporation, 37 percent to radiation, and 42 percent to convection.

TABLE VII

Variation in Percentage of Heat Loss by Radiation, Convection, and Evaporation under Different Conditions

(from DuBois, **14**)

Conditions	Temp. °C	Calories per Hour		Percent Heat Loss		
		Produced	Estimated	Rad.	Conv.	Evap.
Basal, pajamas and sheet	25.0	68.1	76.6	53	19	28
Basal, nude	25.0	64.1	85.0	67	10	23
Exercise, nude	25.0	125.3	105.3	53	25	22
Basal, nude with fan	29.4	68.5	80.4	41	33	20
Basal, nude	34.7	67.3	72.8	5	7	88
Basal, nude with fan	34.7	69.5	79.1	4	6	90

Variations in atmospheric conditions will, however, produce the most diverse ratios. DuBois (**14**) reports figures given in Table VII.

Clearly no general statement as to the percentage relation of various avenues of heat loss can be useful. By means of partitional calorimetry, however, it is not difficult to predict the relations that will obtain in a given set of circumstances.

Changes in Heat Exchange with Varying Environmental Temperature

The fundamental heat phenomena involved in temperature adaptation in the case of man as analyzed by the application of equation 13 are illustrated by such a graph as that presented in Fig. 5. The particular studies concerned included 35 different series of experiments with a single nude subject, observed in the semi-reclining posture. The abscissa represents operative temperatures, and the ordinate represents heat gain *above*

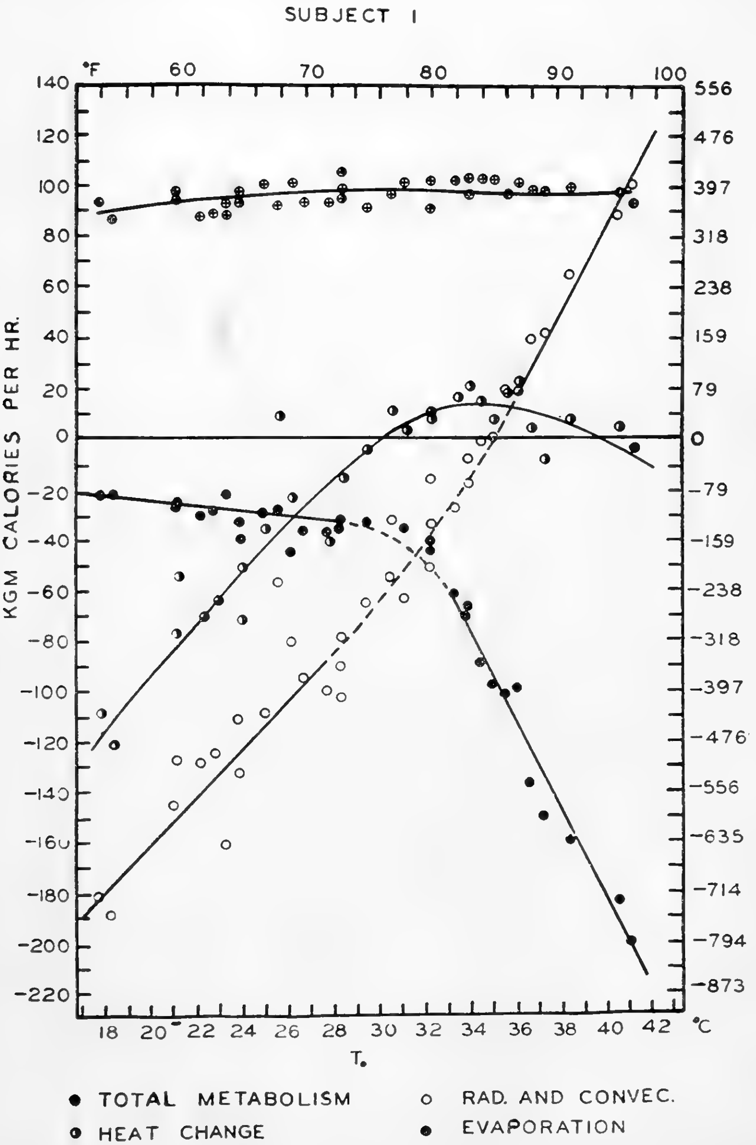

Fig. 5. General diagram of heat loss. Factors in the resting heat balance between the unclothed human body and its environment at various operative temperatures as determined by partitional calorimetry. Ordinal values of lower curves have algebraic sum equivalent to ordinal value of metabolism at a given abscissa value.

the zero line and heat loss *below* the zero line in kg.cal. per man per hour. Each abscissa temperature, if followed upward, will be found to cut four trend lines, and the values on the four trend lines thus intersected represent the average values for the elements of heat loss at that operative temperature.

The upper line of the graph (crossed circles) indicates observed metabolism which, in this case (that of a very stout subject weighing 230 pounds) showed a very slight tendency to increase with rising environmental temperatures. A minimum figure of 87 Calories was observed at operative temperature of 18.5°C, and a maximum value of 106 Calories at a temperature of 28.5°C; both were exceptional. All other values were between 91 and 104 Calories.

The open circles represent the heat interchange by convection and radiation combined, since the abscissa is plotted on the basis of operative temperature which expressed the combined effect of both processes. It will be noted that heat gain, due to these factors (above an operative temperature of 35°C) shows a straight line relationship to operative temperature. Below 35°, however, heat loss due to conduction plus radiation does not fall off as rapidly as operative temperature. The change in slope is due to physiological adaptations in the form of a decrease in skin temperature, which will be discussed in a succeeding paragraph.

The curve for evaporative heat loss (solid circles) shows a gradual increase in the cold zone, from 22 Calories at 18.5° and 23.5°C to 33 Calories at 29.5°C. Above 33°, evaporative heat loss increases very rapidly—to 61 at 33.5° and to 192 at 41.0°C. The slight increase in evaporative cooling between 18° and 30°C is due to the physical factor of changes in atmospheric vapor tension which produces greater evaporation at higher air temperatures. The rapid increase in evaporation above 30°C is due to increased sweat secretion, which will be analyzed in a subsequent paragraph.

Heat exchange is positive (the mean temperature of body tissues is rising) at operative temperatures above 30°C, except in one single instance (with very high wall temperature), where a negative change of 9 Calories was recorded. The positive values varied from 4 to 22 Calories, and the fact that the higher figures appear between 30° and 36°C (with low figures above 36°), may be coincidental or may be due to the very effective evaporative cooling at high temperatures. At operative temperatures below 30°C, values for heat change were, except in a single instance, negative (cooling of body tissues). With minor deviations, the increase in negative heat change is steady from 6 Calories at 29.5°C to 123 Calories and 110 Calories at 18.5°C.

Considering the data of Fig. 5 as a whole, we note three major zones of heat regulation which may be discussed separately.

Zone of Vasomotor Control and Unstressed Evaporative Regulation (*Comfort Zone*)

At operative temperatures between 28° and 33°C (with the particular nude subject in semi-reclining position) there is a zone of essential thermal equilibrium in which *vasomotor regulation* and *unstressed evaporation* are characteristic. This *unconscious* regulation is brought about by minimal sensations from the skin, inducing changes in moisture on the skin and quantity of blood flowing to the skin, in such a way that a perfect heat balance is maintained. In environments either hotter or cooler the large volume of sensation due to stimulation of the temperature receptors of the skin is undoubtedly the source of much discomfort and distraction, thereby causing a reduction in the over-all efficiency of the man. The mean of 36 individual experiments in the comfort zone shows a metabolism of 95 Calories, a heat loss by evaporation of 42 Calories, and a positive heat change of 7 Calories. This state of relative equilibrium is, therefore, characterized by a moderate ratio of evaporative cooling (47 percent of total heat loss), indicating a substantial acceleration of sweat secretion above minimal values. At the lower end of this zone, evaporation is minimal and vasomotor regulation of skin insulation is the primary regulatory process. Near the higher region of the zone, moderate evaporation combines with vasomotor regulation. Ease of regulation is the feature of this region. This area of unstressed equilibrium would, of course, be shifted up or down the operative temperature scale for a subject of different body build or in different positions, or for a clothed subject, or one performing physical work.

The mean values for a group of subjects (**17**) indicate thermal equilibrium between 28° and 33°C, marked negative heat change, and strong skin vasoconstriction, appearing below 28°, and dominant sweat secretion above 33°C. With clothed subjects, on the other hand, the area of equilibrium is wider, as might be expected, and it occurs at a lower environmental temperature. Negative heat change began in this case only below 25°C and active sweat secretion only above 29°C. With a subject performing active physical work (**45**), more variable results were obtained. The highest positive figure for heat change was 83 Calories (with no other positive value over 31); the highest negative value was −47 Calories, although operative temperatures went as low as 12.4°C. These are heat changes per man (not per square meter), and the activity of the working subject corresponded to metabolic rates of over 300 Calories. Figures for evaporative heat loss were also more variable with active work, and they indicated onset of active sweat secretion at 19°–21°C operative temperature. The mean evaporative heat loss for two subjects was 48 Calories at temperatures below 19°C and 152 Calories above 21°C.

Thus, the area of thermal equilibrium may vary from 28°–33°C for the nude subject at rest to 19°–21°C for the same subject performing active work.

Zone of Body Cooling

At operative temperatures below 28°C in Fig. 5 is an area known as the Zone of Body Cooling. As one proceeds downward to successively lower environmental temperatures, heat loss due to radiation plus convection increases progressively, from 46 Calories to observed values of 181 and 188 Calories in these experiments. In the same range, heat loss due to evaporation decreases only from 42 to 22 Calories. The balance is taken up by heat change in the body tissues, which rises from −6 Calories at equilibrium to −116 Calories at operative temperatures of 18° and 18.5°C. The line representing negative heat change is essentially parallel to the line representing heat loss by convection plus radiation.

Zone of Evaporative Regulation

Above operative temperatures of 33°C (for the nude resting subject) lies the Zone of Evaporative Regulation. Here, as one passes to higher and higher environmental temperatures, heat loss due to convection plus radiation decreases and, above an operative temperature of 35°C, changes to a progressively increasing heat gain. While 46 Calories were lost by this avenue in the zone of easy equilibrium, 102 Calories were gained from an environment at an operative temperature of 41°C. However, increased heat loss from evaporation balances this situation with precision, rising from 42 Calories at equilibrium to 192 Calories at 41°C operative temperature. As a result of this adaptive process, there is no appreciable increase in positive heat change. The value of this factor was 6 Calories for equilibrium and 4 Calories for environments at 40°C and over.

In general, the results of the Pierce Laboratory experiments correspond with reasonable closeness to the findings of the earlier workers of the Laboratory of the American Society of Heating and Ventilating Engineers at Pittsburgh. The Heating, Ventilating and Air-Conditioning Guide of the

American Society of Heating and Ventilating Engineers (**2**) compares the two sets of data in a graph which shows general concordance.

Comparison of Nude and Clothed Subjects With Respect to the Temperature Ranges of the Zones of Heat Regulation

The relative reaction of nude and clothed subjects in terms of skin temperature and the nude subjects is given in Fig. 6 as 29°–31°C, which is narrower than that of 28°–33°C quoted in a preceding paragraph. This is not an experimental difference, but reflects the present belief of the authors that the essential characteristic of the middle zone is unstressed thermal equilibrium rather than strict vasomotor control with minimal evaporation.

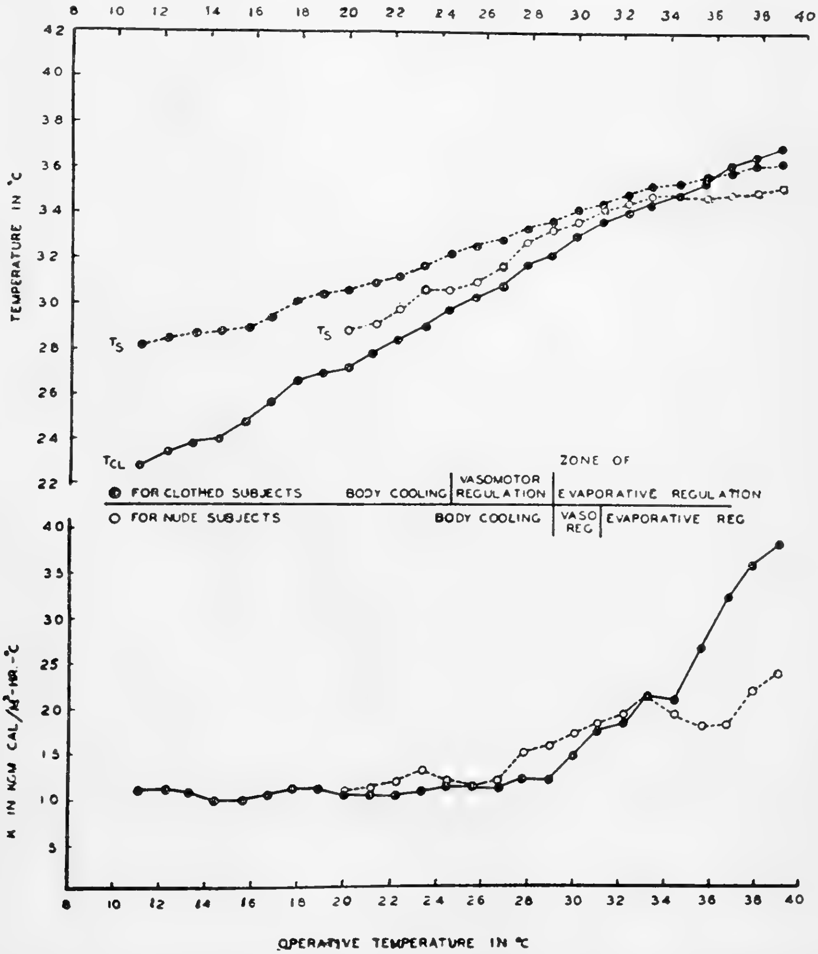

Fig. 6. Comparison of mean skin temperature and conductance (internal regions to skin surface) of resting clothed and nude subjects with respect to the environmental temperatures associated with the three zones of heat regulation.

different phases of regulation for different operative temperatures may be compared in Fig. 6, which is based on data reported by Gagge, Winslow, and Herrington (**17**). The range of the zone of vasomotor regulation for

In general the effect of clothing is to broaden the central zone of vasomotor control and moderate sweat secretion, and to reduce the cooling effect below this zone and exaggerate the heating effect in the region above the central zone. If a single point in the middle of the range is considered, the addition of normal clothing permits a heat balance near 22°C which is similar in comfort and body temperature values to that seen in nude subjects near 28°C.

Reactions of the Body to a Cold Environment

In the Zone of Body Cooling, the primary reaction of the body is a lowering of skin temperature due to constriction of the superficial blood vessels of the skin. Under conditions of long exposure to a chilling atmosphere, there would be, as noted above, a compensatory rise in metabolism, generally associated with shivering and with a tendency to move about.

TABLE VIII

AVERAGE THERMAL CONDUCTANCE OF TISSUES FOR THE NUDE SUBJECT ABOVE AND BELOW 30°C

Temperature °C	Conductance* K		
	Minimum	Mean	Maximum
Above 30°..............	18.0	20.3	23.8
Below 30°..............	10.6	12.6	15.9

* In kg.cal./m.²/hr./°C.

The influence of the fall in skin temperature upon heat loss may be analyzed in terms of a factor referred to as skin conductance, which is included in Fig. 6. The total heat flow from the interior of the body to its surface is equal to the difference between metabolism and positive heat change, or to the sum of metabolism and negative heat change. If we divide this figure by the surface area of the body and divide again by the difference between rectal temperature and skin temperature, we have a measure of heat flux per unit area of skin. In other words,

$$(14) \qquad \frac{M - \Delta H}{A} \div (T_R - T_S) = K.$$

The value of conductance (K), determined in this way, shows (for the same nude subject for whom data are presented in Fig. 6) the differences, above and below the critical temperature given in Table VIII.

The conductance, as thus recorded, obviously depends on two major factors, thermal conductance of the body tissues, and the flow of blood which brings warm blood to the surface to be cooled. The first of these factors is influenced, however, not only by the specific conductivity of the flesh itself but also by the depth of surface tissues that are actually cooled below a normal value. This gradient is itself obviously influenced to a considerable extent by blood flow, but is also affected by the progressive chilling of deeper and deeper layers of body tissue. The fact that the rate of lowering of skin temperature (and hence of conductance) at low operative temperatures is so much more rapid than the rate of rise in skin temperature at high operative temperatures indicates clearly the importance of vasoconstriction; and the suddenness with which the picture changes near 30°C for the nude subject suggests a definite reflex response.

It should be emphasized that quantitative values may vary widely in different subjects. While a very stout subject showed a conductance of 18.3 under warm conditions and a conductance of 12.4 under cold conditions, the corresponding values for a less well-insulated slender subject were 25.0 and 13.7. The second subject, too, showed very high values (K over 35) at extremely high operative temperatures. In both cases, however, there was the same sharp break near 30°C operative temperature, indicative of vasoconstriction.

The physiological reactions in this process have been reviewed in some detail elsewhere (**17**) with reference to the extensive studies of Kleiber, Burton, Bazett, Hardy, and DuBois. All that need be said further here is that below an operative temperature of 28°C, for both nude and clothed subjects, the value of K lies approximately constant between 10 and 12 (with gradients between rectal and skin temperature varying from

4.5° up to 9.5°C, as one proceeds to cooler and cooler environments). At operative temperatures over 30°C, the *K* values rise more or less steadily with higher operative temperatures, reaching 23.8 for the nude and 38.7 for the clothed subjects at an operative temperature of 38.9°C. In this area the great increase in conductance is accompanied by progressive reduction in the gradient between rectal and skin temperature (4.5°–1.5°C). Clearly, in the Zone of Body Cooling, a maximum vasoconstriction has been reached, and *K* remains essentially constant.

layers of the body equal to 21 percent of the normal specific output of the heart.

The lowering of conductance by vasoconstriction in cold environments, together with the lowering of skin temperature which accompanies it, is a useful adaptative mechanism for reducing the strain put upon the body by such environments. That this mechanism is quite unable to cope effectively with any considerable fall in operative temperature is clear, however, from the rapid increase in negative heat change that is noted below 28°C operative temperature

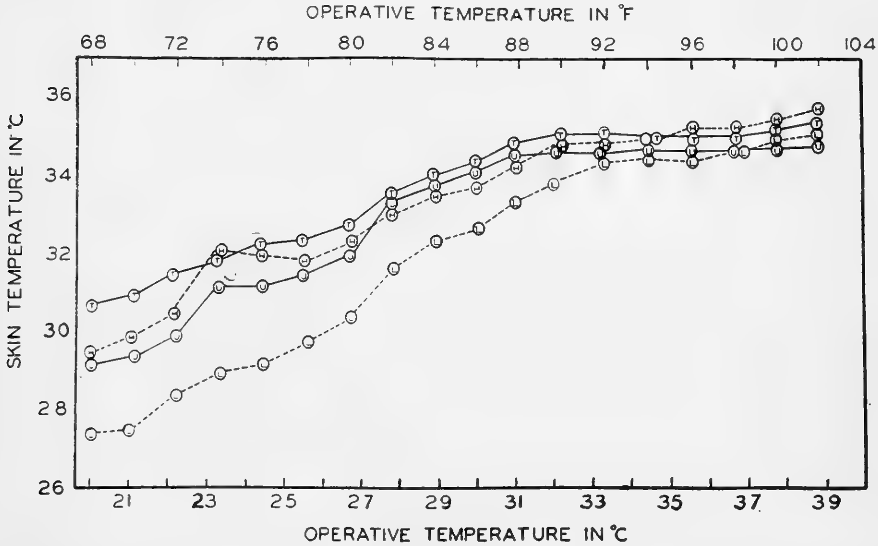

Fig. 7. Mean values for the segmental skin temperatures of four unclothed subjects at operative temperatures from 20° to 39° C. T, H, U, L, represent Trunk, Head, Upper, and Lower Extremities.

In the Zone of Evaporative Regulation, on the other hand, the relatively rapid rise of conductance must be due chiefly to progressively increasing blood flow. Analysis of the data indicates that, under the critical conditions where there is a minimum *K* value and a difference of 4.5°C between skin and body temperature, the depth of the tissue layer cooled below the general body temperature is of the order of 2.2 cm. The increase in blood flow corresponding to the conductance of 38.9 observed at 39°C operative temperature for the clothed subject indicates an increase in blood volume in the peripheral

with the nude subject, and below 20°C with the clothed subject. Clearly, such marked cooling of body tissues could not continue progressively for any extended period without serious interference with bodily functions.

Actual Skin Temperature for Various Regions of the Body Surface Beyond the Area of Equilibrium

Fig. 7 shows the difference in skin temperature for four of the main regions of the body at 20 different operative temperatures, using the mean values for four nude subjects in the usual semi-reclining position. It will be noted that in the Zone of Evaporative Regulation (at the right of the chart) values for the trunk remain constant (between 34.9° and 35.2°C), and values for the upper extremities also remain constant (34.6°–34.7°C). Values for the head rise slightly with rising environmental temperature in this Zone (from 34.8°C at operative temperature 32.2° to 35.7° at operative temperature 38.9°C). The temperature of the legs also rises from 34.4° at operative temperature of 33.3° to 35.1°C at operative temperature of 38.9°C.

Below 33°C operative temperature for the lower extremities, and below 32°C operative temperature for the other three areas, all skin temperatures begin to fall more rapidly as operative temperatures fall (passing toward the left of the chart). The rate of fall is least rapid with the trunk and most rapid with the lower extremities.

Physiological Adaptations in the Zone of Evaporative Regulation

In this region, which is above 33°C for the nude and 29°C for the clothed individual, the body has an extraordinarily effective process of compensation for an environment of high heat stress, a compensation which, in its physical manifestations, is very simple, being accomplished by a progressive increase in evaporation. The perfection of this reaction, however, involves physiological adjustments of considerable complexity.

Under the conditions of ordinary studies of basal metabolism, with reasonably standard conditions of thermal environment, the evaporative heat loss from the body is moderately low and highly constant. DuBois (**13**) tabulates a large number of experiments of this type for normal persons and those suffering from various diseases, in which the percentage of metabolic heat lost by evaporation varies only between 21 percent and 28 percent, averaging about 25 percent. Under such conditions, the insensible perspiration bears a direct straight-line relationship to metabolism. Under hot environmental conditions, on the other hand, the evaporative heat loss may rise, as we have seen, to very high values. The obvious environmental factor which is related to the phenomenon of increased sweat secretion is air temperature; but this relationship is only manifest above the critical level of the area of thermal equilibrium. Below an operative temperature of about 30°C, with the nude semi-reclining subject, evaporative heat loss is slight. Above this point comes a sharp increase. This increase in evaporation is, of course, generally proportionate to progressive increase in air and wall temperature; but it is not governed by physical moisture demands of the atmosphere, since it bears no relation to the relative humidity of the atmosphere in this range. The physical capacity of the air to take up moisture changes enormously at various humidities, as indicated in Table IX.

TABLE IX

AMOUNT OF WATER THAT CAN BE RETAINED BY AIR AT VARIOUS TEMPERATURES

Temperature °C	Grams per cubic meter
−20	.9
−10	2.1
0	4.8
10	9.3
20	17.0
30	29.9

Air which is completely saturated at 0°C is only 28 percent saturated when warmed to 20°C. Yet, in spite of wide variation in temperature and relative humidity, the actual evaporation from the body over a very considerable thermal range is maintained at equilibrium level by a varying rate of secretion of sweat on the body surface in accord with specific adaptive physiological processes (**43**).

The best index of the rate of evaporative heat loss is to be found in skin temperature. Evaporative heat loss remains fairly constant (below 20 Calories per square meter) up to a mean skin temperature of 32°C. Above this point it rises progressively to over 90 Calories per square meter at a skin temperature of over 35°C. Thus, while evaporation remains relatively constant in the Zone of Body Cooling as skin temperature falls, the reverse is the case in the Zone of Evaporative Regulation.

In analyzing the process of evaporative regulation, it has been convenient to introduce a new physiological concept, that of wetted area (**15**). This factor is discussed in the section on Physical Processes under Evaporation. The utility of the wetted area concept is that it permits an estimate of the maximum value of evaporation under stress conditions, and affords a reasonable explanation of the independence of relative humidity and unstressed evaporative heat loss. A further discussion would not be valuable here, except to note that it has been a useful concept in developing the practical values for the upper limits of evaporative regulation.

The Upper Limits of Evaporative Regulation

Precise and effective as the process of evaporative regulation is, it operates only within certain defined limits. The most important limiting factor is the amount of moisture that the atmosphere can actually absorb from a saturated surface at skin temperature (35–36°C). If we assume a maximum wetted area and a skin temperature of 35.6°, we can compute for any combination of atmospheric temperature and humidity the number of Calories per unit area of skin which will be evaporated. Knowing the constants for convection and radiation, as determined by methods outlined previously, and assuming a metabolic rate of 47 kg.cal./m.^{2}/hr., we can compute the amount of heat which must be given off by evaporation to maintain equilibrium. Any combination of atmospheric temperature and humidity for which the possible heat loss by evaporation does not equal the evaporative heat loss necessary for equilibrium lies beyond the upper limit of tolerance. With saturated air at any temperature near skin temperature, evaporation is, of course, nearly zero, in accordance with the very small differential in vapor pressure represented by the difference in the saturation values for air and mean skin temperatures.

By the use of this technique of estimating evaporative limits, it has been possible to compute the upper limits of evaporative regulation given in Fig. 8.

The accuracy of such limits is dependent upon the validity of the constants used, which in this case are drawn from the study by Gagge (**15**). For conditions of low air movement, the values are probably quite reliable. The effects of added air movement on wetted area are less well established and the limits noted may be offered as best approximations on the basis of presently available data. The values for increased air movement assume that air movement affects evaporation in a manner analogous to the effect of increased air movement on convection loss.

In a very hot and dry atmosphere, such as one finds in some desert regions, another factor comes into play: the physiological capacity of the body to produce sweat. In such an atmosphere, the difference of vapor pressure between skin and air is so great that, even with maximum vasodilation, the sweat glands are unable to maintain a high wetted area. However, under sub-surface marine conditions this is not the typical case. The three curves of Fig. 8 indicate the upper limits of tolerance fixed by the evaporative power of the atmosphere for nude subjects under three conditions. Curve A is for a subject at work (with a total heat production of 425 Calories per hour and minimal air movement); B is for a subject at rest with a total metabolism of 85 Calories at minimal air movement; C is for a resting subject exposed to air movement at a rate of 100 feet per minute. The crosses on Curves A and C indicate the upper limit of tolerance which would be set by the physiological capacity of the body to produce certain stated amounts of sweat (500, 700, and 900 grams per hour).

At a temperature of 27°C both resting and working subjects can adjust in still air even at 100 percent relative humidity. The resting subject can also adjust with any relative humidity at 32°C and the resting subject in rapidly moving air (100 feet per minute) can adjust with any humidity up to 35°C. The general slope of the lines below these points indicates the limits of possible adjustment based on the absorptive capacity of the atmosphere. The resting nude can, of course, adapt to much more extreme conditions than the working subject, and the zone of adaptation is further broadened by a high degree of air movement (100 feet per minute).

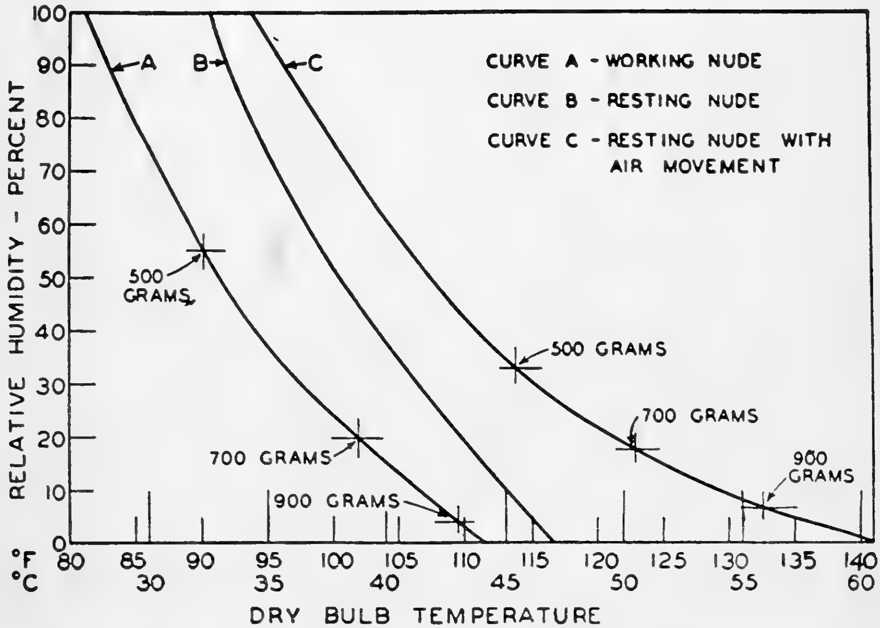

Fig. 8. Upper limits of evaporative regulation. Curve A, working; heat production 425 Cal. per hour, air movement 15 fpm; Curve B, resting; 85 Cal. per hour, air movement 15 fpm: Curve C, resting; 85 Cal. per hour, air movement 100 fpm. Crosses indicate points on limits which can be reached through the indicated evaporative water loss in grams per hour.

In Line A of Fig. 8, it will be noted that at 32°C the upper limit for the working subject is at 55 percent relative humidity; if this line were continued downward and to the right on a basis of atmospheric absorptive power, the working subject would adjust to 39°C and 20 percent relative humidity, and to 43°C and 3 percent relative humidity. In this area, however, the other limiting factor, the power of the body to produce sweat, comes into the picture. In this respect individuals vary widely; and the same individual may greatly increase his ability to secrete sweat. Dill (**11**) has shown that a subject working in hot dry desert air can double his capacity for moisture production, by practice, from 600 grams to 1200 grams per hour (from a little over a pint of water to a little more than a quart). In football games even higher figures have been recorded. If, however, we assume a representative figure for the normal individual, approximating Dill's initial value, it is found that above 32°C the working subject would not be able to produce the amount of moisture necessary for adjustment, even though the atmosphere could absorb it. For the working subject producing 700 grams of moisture, this limit of secreting power would come at 39°C, and for the subject secreting 900 grams at 43°C. Vertical lines drawn downward on the graph would indicate these limits, and similar limits are set for the resting nude in moving air, for the subject producing 500 grams at 46°C, for the subject producing 900 grams at 51°C, and for the subject producing 900 grams at 56°C. At any point below the slope of the graph and to the left of the appropriate vertical line, the subject can maintain thermal balance. Above the slope and to the right of the vertical line, adaptation is no longer possible on the basis of these assumptions.

The general topic of adjustment to extreme heat and cold has been the subject of many excellent studies within the past few years. The relations noted above have been investigated for a number of different conditions of exercise, clothing, and type of climatic exposure. Many of these studies grew out of military interests of the recent war period. This has a bearing on the use of such concepts as limit and tolerance in relation to human climatic endurance. Neither term is susceptible of exact definition, since human motivation inevitably enters in the practical testing of such concepts. In general, the limits suggested by our work as reported in the foregoing discussions are conservative. Interested readers will find in a recent book by Adolph (**1**) an extended discussion of the many factors involved in the problems of limit and tolerance in relation to heat stress. Taylor and Marbarger (**38**) have extended this work to include a special treatment of tolerance times in environments in which ultimate adjustment is not possible.

When the upper limits of evaporative regulation are reached, disastrous consequences ensue. With our clothed subjects, profound discomfort was manifest with high relative humidity (70–80 percent) even at atmospheres between 37° and 39°C, and some subjects showed nausea and other subjective symptoms so severe that the experiments had to be discontinued. Under such conditions, skin temperatures and rectal temperatures begin to rise sharply and, what is even more serious, metabolism rises also, thus creating a vicious spiral.

The data obtained in our Laboratory correspond well with the findings of McConnell, Houghten, and Yaglou (**30**) that saturated air at 32.2°C is the "upper limit of man's ability to compensate for atmospheric conditions" in still air. They are also in accord with the reports of Cadman and Haldane (**7**). Cadman states that at 29.4°C wet-bulb temperature, the body temperature invariably rises, while at 33.9° wet-bulb "one is in a terrible state"; Haldane states that at 31°–32°C wet-bulb "in fairly still air, the body temperature begins to rise, even in the case of persons stripped to the waist and doing no work; and when air is saturated this rise continues until symptoms of heat stroke arise." These limits may be tolerated and exceeded by highly motivated, physically fit, and well-acclimatized subjects. However, no situation is easily conceivable in which the physiological and psychological cost of adjustment can be considered permissible as a standard practice for personnel entrusted with serious duties.

These effects of heat stress are not, of course, instantaneous. The Pittsburgh investigators found exceptional subjects who could endure such extreme conditions as 70°C with a relative humidity of 15 percent for half an hour. The body temperature under such conditions rose several degrees. The pulse rate seemed to be the best measure of sensations, 135 beats per minute corresponding to marked discomfort, and pulse rates exceeding 150 being almost unbearable. Restlessness, irritability, headache, palpitation of the heart, soreness of the eyes, a severe oppression of the chest, dizziness, and confusion, were among the sensations experienced; and weakness and a dragged-out feeling persisted for some time after such experiments had closed.

Under such conditions of very active sweating, the body, of course, suffers from water loss and, in even more serious degree, from loss of sodium chloride. It is, therefore, the custom to provide salt tablets for men who do heavy work under such conditions. The same procedure was followed during the war with troops in the tropics.

In the early days of steam, temperatures in the furnace rooms of naval vessels in the tropics sometimes passed 65°C, with high relative humidity. In 1909 and 1910 the attack rate for heat stroke in the Navy was about 8 per 1000, and between 1861 and 1911, there were 20 deaths and 33 men invalided from this cause. Severe cramps and muscle twitchings attributed to local drying-out of the tissues and accumulation of waste products of metabolism are among the pronounced symptoms of heat stroke, and the body temperature may be so completely upset that temperatures of 40° to 44°C have been recorded, associated, of course, with profound hazard to life.

TABLE X

Mean Rectal Temperatures of Clothed Subjects after 4–8 Hours Exposure to Room Temperatures Between 20°–30°C

Air Temperature	20°	24°	30°
Relative Humidity	50%	50%	80%
Rectal Temperature	36.7°	37.0°	37.4°

Relation of Thermal Conditions to Human Comfort

Even within the range of reasonably effective thermal adaptation, and long before extreme strain occurs, there are certain definite physiological changes which may have significant effects on human comfort. The basic studies of the New York State Commission nearly a quarter of a century ago yielded the following results.

The rectal temperature, when subjects came to the laboratory in the morning, showed a definite relation to the outdoor temperature of the preceding night in the summer season (but not under the artificial conditions of winter life in heated houses). The mean rectal temperature, after 4–8 hours exposure to experimental conditions (subjects normally clothed with moderate activity) is given in Table X.

No difference in blood pressure was noted among the three atmospheric conditions defined above. A special experiment at 38°C with 87 percent relative humidity showed a definite increase in systolic pressure. A comparison of changes in diastolic pressure with changes in the product of the pulse rate showed that the resistance in the peripheral portion of the circulatory system was decreased in the warmer conditions.

Particularly interesting results were obtained with regard to a factor known as the Crampton Index, which is one of many relative measures of the efficiency of cardiac and circulatory adjustments to various stresses. This index is based on the ratio between increase in heart rate and rise or fall of blood pressure when the subject passes from a reclining to a standing position. It measures, in an approximate manner, the effectiveness and economy of adaptation to the extra burden of erect posture and is presumably a function of vasomotor reaction. The scale on which the Crampton Index is expressed is so designed that a figure of 100 corresponds to an increase of 10 millimeters of blood pressure with an increase of less than 4 beats in the heart rate; a value of zero corresponds to a decrease of 10 millimeters of blood pressure with an increase of over 40 in the heart beat. The values given in Table XI were recorded in the Commission experiments. Higher values of the index noted at 20°C are considered as indications of better circulatory adjustment to the gravity stress of the erect posture. Since such substantial

TABLE XI

Changes in the Crampton Index of Vasomotor Response to Postural Change in Clothed Subjects at Temperatures Between 20°–30°C

Air Temperature	20°	24°	30°
Relative Humidity	50%	50%	80%
Crampton Index	60	45	34

variations in physiological status occur, it is to be expected that hot conditions, well within the range of thermal equilibrium, may be accompanied by sensations of discomfort.

The fundamental studies on this latter point were made at the Research Laboratories of the American Society of Heating and Ventilating Engineers (to which full reference will be found in the current issue of their *Guide* (**2**). These tests were based primarily on indications of preference expressed by subjects passing from one carefully regulated room to another room with slightly different temperature or humidity. The major result of these studies was the development of what is known as the Effective Temperature Index. This index was determined for clothed subjects not engaged in active physical work and exposed to minimal air movement (15–25 feet per minute). It represents, for these conditions, the net effect upon comfort of variations in atmospheric temperature and humidity. The Effective Temperature is defined as that temperature of completely saturated air which will produce the same subjective sensation of comfort as the particular combination of temperature and humidity observed (in both cases with minimal air movement). Similar charts have been drawn for higher air velocities; and the way in which the scale works is illustrated in Table XII, showing various combinations of air conditions, all producing an Effective Temperature of 17.2°C.

Fig. 9, again reproduced by the courtesy of the American Society of Heating and Ventilating Engineers, shows the zones of dry-bulb and wet-bulb temperatures that correspond to maximum comfort as determined in the Pittsburgh studies. It will be noted that two separate zones are indicated for winter and for summer comfort, respectively, since it was found that the subjects in the summer preferred distinctly higher temperatures than in winter. The winter comfort zone extends from an Effective Temperature of 63°F (17.2°C) to one of 71°F (21.7°C). The summer comfort zone ranges between 66°F (18.9°C) and 75°F (23.9°C). An optimal area for all seasons is indicated between 66°F and 71°F (18.9°C and 21.7°C) with relative humidities between 30 and 70 percent.

The Effective Temperature data were obtained in rooms with air and walls at approximately the same temperature. Later A.S.H.V.E. studies have led to the conclusion that an elevation or lowering of the mean radiant temperature of the enclosure 1°F above or below the air temperature can be balanced by a 0.5°E.T. lowering or raising of Effective Temperature. The correction for mean radiant temperature is obviously a very rough approximation. The relative

TABLE XII

Air Temperatures, °C, Producing Equivalent Sensations of Comfort, Equal to an Effective Temperature of 17.2°C, with Varying Degrees of Relative Humidity and Air Movement

Relative Humidity %	Rate of Air Movement, Feet per Minute			
	Minimal	100	300	700
20	20.2	21.4	23.3	25.0
50	19.2	20.3	22.5	24.2
100	17.2	18.6	20.8	22.8

influence of air and wall temperatures can only be evaluated by the factor of Operative Temperature discussed earlier. It would be much sounder to determine operative temperature directly and then apply it, in place of air temperature, in the Comfort Chart.

A.S.H.V.E. studies (**2**) on both resting and working subjects have shown the striking effects of High Effective Temperature on rectal temperature and pulse rate as given in Table XIII. This table affords ample evidence of the stress imposed by saturated environments (or their equivalent) above 32.2°C (90°F) on either resting or working subjects. From the standpoint of service engineers, the details of this stress are less important than a clear realization of the reality of the severe physiological stress that may be eliminated by air conditioning equipment able to maintain quarters well below the stress conditions of Table XIII.

It should be mentioned here that the A.S.H.V.E. Effective Temperature Scale is proximately 27°C (81°F). In the lower comfort range, the scale likewise serves to equate the equivalent sensation effects of humidity and temperature for contrast situations in which a person passes from a moder-

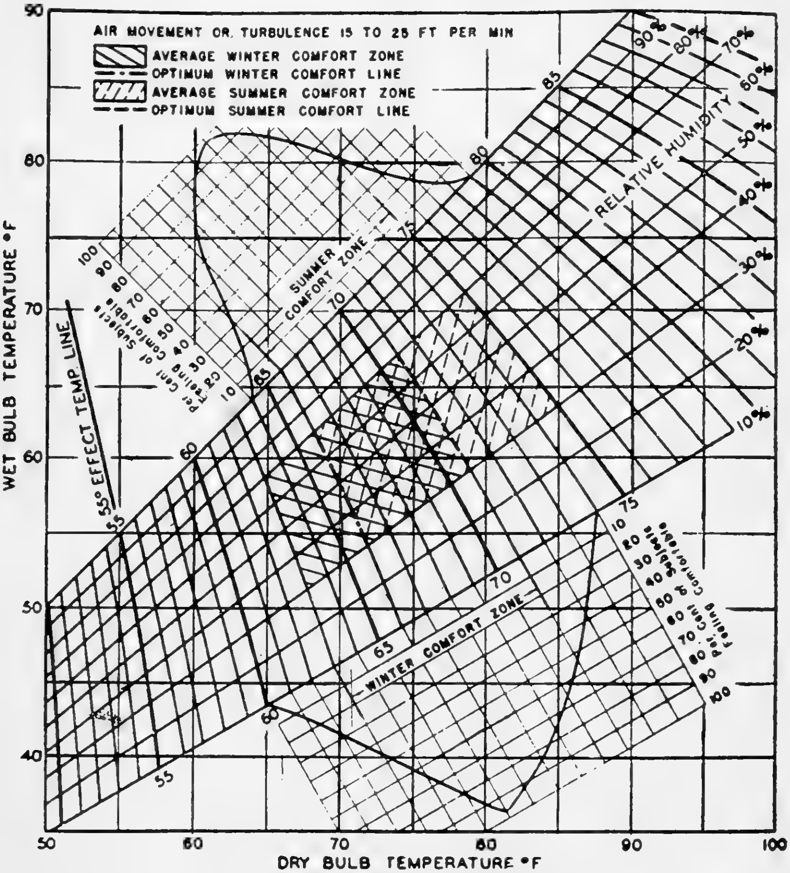

Fig. 9. American society of heating and ventilating engineers' comfort chart for still air. Both summer and winter comfort zones apply to inhabitants of United States only. Application of winter comfort line is further limited to rooms heated by central station systems of the convection type. The line does not apply to rooms heated by radiant methods. Application of summer comfort line is limited to homes, offices, and the like, where the occupants become fully adapted to the artificial air conditions. The line does not apply to theatres, department stores, and the like, where the exposure is less than three hours. The optimum summer comfort line shown pertains to Pittsburgh and to the other cities in the northern portion of the United States and Southern Canada, and at elevations not in excess of 1000 feet above sea level. An increase of 1° ET should be made approximately per 5° reduction in north latitude.

(Reproduced by courtesy of the American Society of Heating and Ventilating Engineers)

in process of being revised and extended in scope at the present. It has been realized for some time that the scale provides a very satisfactory index of equivalent condition at dry-bulb temperatures above apate dry-bulb atmosphere to a similar temperature at high (or lower) humidity. After adaptation over a period of one to two hours, however, the sensations realized are not in complete accordance with the scale. In this moderate temperature range (20°–27°C, 68°–81°F), the effect of humidity is overemphasized.

TABLE XIII

THE STRESS EFFECT OF RISING EFFECTIVE TEMPERATURE ON CIRCULATION AND WATER BALANCE

(American Society of Heating and Ventilating Engineers, **2**)

Effective Temp. °C	Men at Rest			Men at Work, 90,000 f.p./hr.		
	Rise in Rectal Temp./hr. °C	Increase in Pulse Rate, Beats/min. per hr.	Approx. Sweat Loss kg./hr.	Rise in Rectal Temp./hr. °C	Increase in Pulse Rate, Beats/min. per hr.	Approx. Sweat Loss kg./hr.
29.4	.06	1	.18	.33	17	.50
32.2	.17	4	.23	.66	31	.68
35.0	.51	15	.41	1.3	61	.91
37.8	1.2	40	.77	2.2	103*	1.22*
40.6	2.2	83	1.22	3.3	158*	1.59*
43.3	3.3	137*	1.81*	4.7	237*	1.99*

* Computed for exposures lasting less than one hour.

tive humidity in the hot zone. It should be remembered that Fig. 10 applies to nude subjects in a semi-reclining position. The comfort chart of the American Society of Heating and Ventilating Engineers was based on studies with subjects who were allowed a certain freedom of movement and had a substantially higher metabolism. They were fully clothed under cold conditions and stripped to the waist under hot conditions.

Physical Cost of Excess Clothing Under Hot Conditions

Aside from simple over-heating, the principal effects of clothing under warm conditions are (1) a possible reduction of solar heat gain and an increase in evaporative efficiency under certain circumstances, and (2) an increase in circulatory stress and dehydration effects where more clothing is required for other reasons than is desirable from a purely thermal standpoint. In Table XIV, the vari-

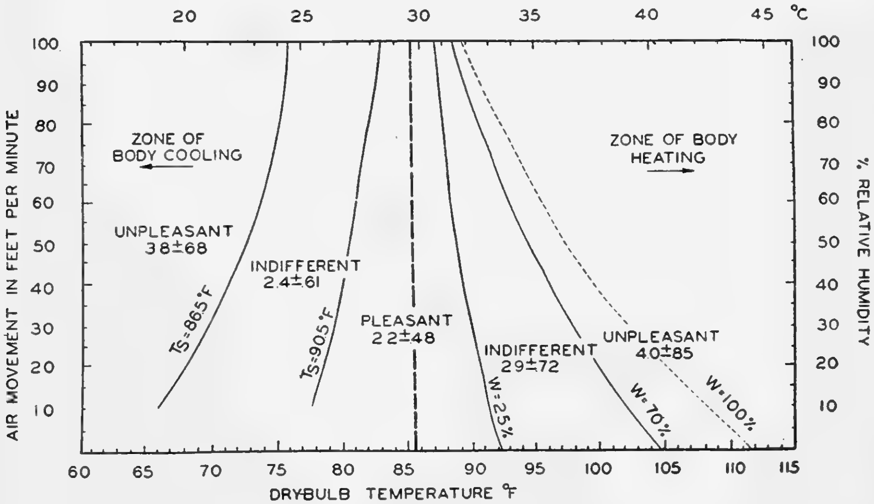

Fig. 10. The relation of air movement, relative humidity, and operative temperature, to thermal comfort as experienced by unclothed subjects at rest

The general relations of sensations of comfort are presented graphically in Fig. 10, plotted on the abscissa against dry-bulb temperature and on the ordinate against air movement in the cool zone and against relative humidity in the hot zone.

able dehydration stress that may be imposed by clothing is illustrated. These results show the sweat secretion of a young male subject working on a bicycle ergometer with a total heat production of 279 kg. cal. per hr. This work is approximately equal to walking at three miles per hour with a 20-pound pack. These data show that at air temperatures above body temperature, with low humidity, the effect of clothing variation is not conspicuous at low air velocities. However, at moderately hot temperatures with high humidity, the dehydration effect of added clothing is conspicuous. The thermal analysis of this differential effect is complicated, but, in general, the difference depends upon the importance of radiation and convection cooling at the 29.4°C (85°F) condition and the dominance of evaporative cooling at 43.3°C (110°F).

TABLE XIV

DEHYDRATION STRESS RELATED TO CLOTHING IN SUBJECTS WORKING UNDER HOT CONDITIONS (TOTAL METABOLISM, 279 KG.CAL./HR.)

Dress	Sweat Secretion in grams/hour	
	at 29.4°C (85.0°F) 85% R.H.	at 43.3°C (110°F) 15% R.H.
	(equivalent air movement 100 ft./min.)	
Light underwear, shorts, moderate weight pants, and shirt, shoes, light socks	470	700
Stripped to the waist	438	603
Shorts, underwear, socks, shoes	333	638
Athletic supporter, socks and shoes	260	652

In Table XV are summarized data from Yaglou (**47**), relating clothing, comfort sensations, and body temperatures, for individuals under sedentary conditions. As noted earlier, work of the order of three times this resting level of heat production reduces the comfort level for the nude subject by 12°C (18°F). Hence, for partially stripped working subjects, we may approximate the comfort point at work by allowing a decrease of 4°C (7.2°F) from the level of sedentary comfort for each increase of 100 percent in heat production over the basal level.

The Clo Concept and Unit of Clothing Insulation

Herrington (**27**) has discussed the joint influence of clothing and level of heat production on thermal adaptation, and gives particular mention to the "clo" concept of clothing insulation. The widespread use of partitional methods of estimating heat and cold stress from subjects and models has created a need for a unit of heat insulation which would have practical meaning for nontechnical groups. Such a unit should be convertible into the BTU/sq.ft./°F used by

TABLE XV

EFFECT OF VARIATION IN CLOTHING ON THE TEMPERATURE OF THERMAL COMFORT AT REST (from Yaglou, **47**)

	Clothing Weight		Mean Temperature of Comfort								Metabolic Rate kg. cal./hr.	No. of Subj.
			Air		Skin		Clothing		Rectal			
	kg.	lb.	°C	°F	°C	°F	°C	°F	°C	°F		
Men unclothed	0	0	28.9	84.0	33.9	93.1	—	—	37.0	98.6	81	2
Women's summer clothing	.8	1.8	26.7	80.0	33.5	92.3	31.8	89.3	37.0	98.7	74	2
Men's summer clothing	2.2	4.9	24.4	76.0	34.0	93.2	30.1	86.1	37.2	98.9	79	17
Men's indoor winter clothing	3.8	8.4	21.9	71.5	33.3	92.0	28.1	82.6	37.0	98.6	79	18
Men's outdoor winter clothing	7.3	16.2	11.9	53.5	33.2	91.8	19.8	67.6	37.3	99.1	83	2

the ventilating engineers, the physicist's gm.cal./sec./°C, and the physiologist's kg. cal./hr./°C, without being dependent upon these units for an approximate subjective appreciation of its insulation value. With these points in mind, Gagge, Burton, and Bazett (**16**) defined a practical unit, the "clo." One clo unit of thermal insulation is the clothing required to keep a resting subject in a comfortable state when the subject is seated in an atmosphere of 21.1°C (70°F) with relative humidity less than 50 percent and air movement at 20 ft./min.(10 cm./sec.). The standard value for the metabolism associated with this condition is 50 kg.cal./m_2./hr.

The standardization of the clo unit utilized the experimental work of Winslow, Gagge, and Herrington (**44**) to determine that the total insulation, which is the sum of the insulation of the clothing, I_{Cl}, and of the air, I_A, is

$$I_{Cl} + I_A = \frac{33\text{–}21}{38} = 0.32 \frac{°\text{C}}{\text{kg. cal./hr./m.}^2}. \tag{15}$$

The insulation of the air in metric units at the air movement cited is

$$0.14 \frac{°\text{C}}{\text{kg. cal./hr./m.}^2}. \tag{16}$$

By difference, the insulation of the clothing, equal in the above definition to one clo, is

$$0.32 - 0.14 = 0.18 \frac{°\text{C}}{\text{kg. cal./hr./m.}^2}, \tag{17}$$

or

$$= 0.88 \frac{°\text{C}}{\text{BTU/hr./ft.}^2}.$$

The clo definition is in terms of resistance, rather than conductance, units, since the former may be added directly to obtain the total resistance from known sub-components. Perhaps something is added to the sensory appreciation of this unit by the following example. With low air movement and moderate humidity, the resting nude subject is comfortable at 30°C (86°F). One clo of insulation is required to maintain the same degree of comfort when the air temperature is dropped 8.8°C (16°F). Extending this analogy, one can say that two clo is the clothing sufficient for comfort at 30°–17.6°C or 12.4°C (54°F), three clo that sufficient for 30°–26.4°C or 3.6°C (38°F), and so on, it being understood that air movement and heat production remain in agreement with the formal definition.

It is probably obvious that one clo is very nearly the insulation provided by normal male clothing. This is intentional, and provides a valuable experience correlation for the unit.

In Fig. 11 five illustrative pictograms and graphs have been redrawn from an article by Burton (**8**) on clothing in relation to human heat exchange. Presentation of this type has been useful in demonstrating to lay authorities the practical significance of the clo unit of clothing insulation in relation to the met unit of human heat production under varied climatic exposures. In the pictograms labeled 1, 3, and 6 met, the position of the mercury at +70°F (21.1°C), +40°F (4.4°C), and −5°F (−20.6°C) illustrates the approximate decrease in environmental temperature which may be tolerated with standard clothing as heat production is increased from a sedentary level of one met (50 Cal./m^2./hr.) to six met, which represents severe exertion. The fourth exercising figure indicates that an increase of clothing insulation to $1\frac{1}{2}$ clo with a heat production of six met permits a further drop in environmental temperature to −30°F (−34.4°C). The two sleeping-bag pictograms illustrate the choice that may be made between heavy or medium protection in sleeping bags (11 or 8 clo insulation value), provided that auxiliary heating (stove) is used to raise a shelter temperature of −20°F (−28.9°C) to +10°F (−12.2°C). In the outline figure of a clothed man, a graphic indication has been given of the three components into which the total insulation of the clothed man has been analyzed by Burton (8).

The mitten illustration emphasizes the difficulties which are met in trying to provide adequate insulation for the hands at an environmental temperature of −20°F (−28.9°C). With this exposure, at rest, the amount of insulation required by the hand is greatly in excess of what is practically possible if the hand is to remain functionally useful. If the time of exposure is limited to two to three hours, protection may be provided with gloves whose bulk is within practical limits. Finally, if exercise is strenuous, the glove may be dispensed with. The mitten illustration and the graph in the lower

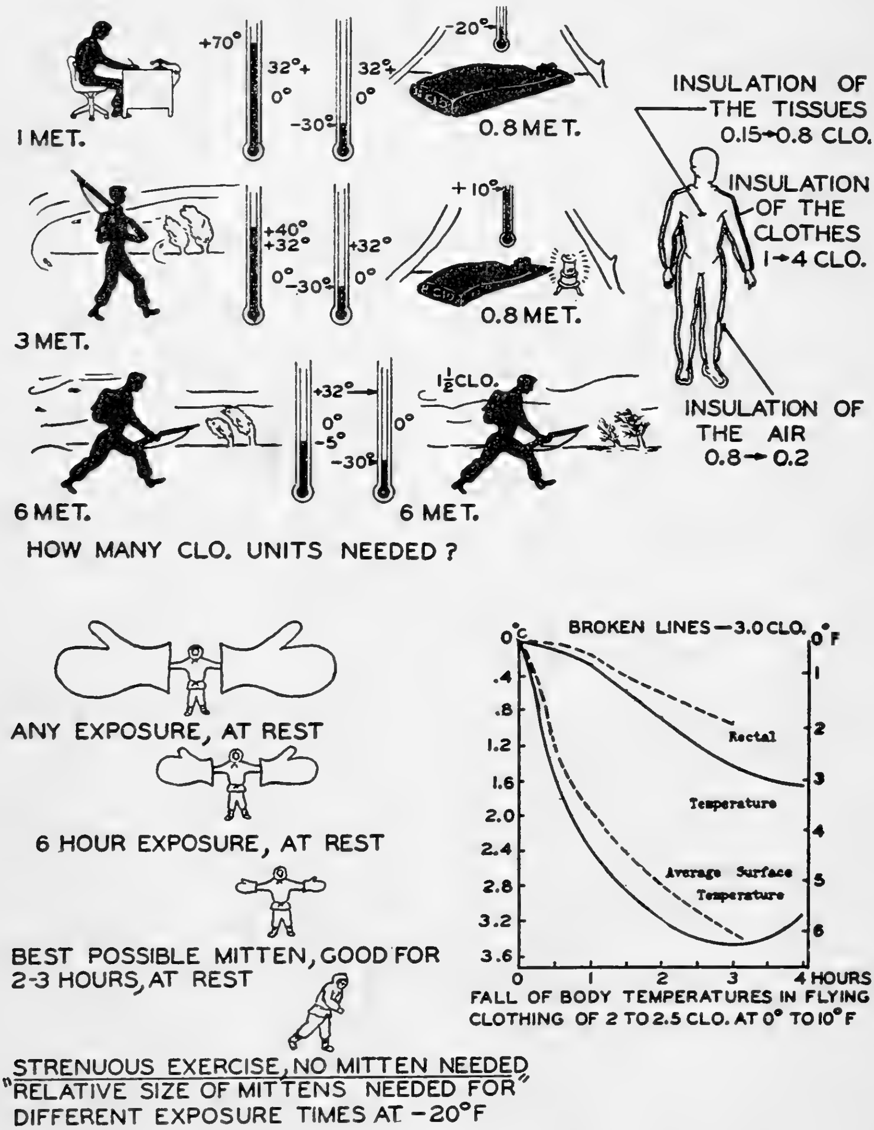

Fig. 11. Graphic illustration of the interdependence of level of heat production, grade of insulation, and severity of exposure on the adequacy of protective clothing
(Redrawn from A. C. Burton, 8)

right hand portion of Fig. 11 were based on studies at the Army Climatic Laboratory. The graph illustrates the fall in rectal and skin temperature observed in 50 subjects seated in a cold chamber and clothed in heavy flying clothing. Apparently, under heavily clothed conditions the uniform protection of the skin tends to delay the onset of protective shivering (increased heat production), and lower body temperatures occur than are seen when men more lightly clothed are exposed to limited periods of cold stress.

Climatic maps have been prepared which relate the clo unit of insulation to various climates of the world, and the concept has been generally used by the majority of investigators who have attempted a correlation of scientific information with practical exposure problems.

Comment Relative to Temperature and Humidity Conditions in Undersea Operations

The foregoing material is intended to be informative rather than minutely prescriptive with reference to principles governing thermal hygiene, since the many circumstances that limit the achievement of ideal air conditions in undersea operations are well known. A survey of operational experience indicates that air-conditioning systems are absolutely essential for effective undersea operations. Present systems are only moderately effective. It is believed that the redesign of such systems in a manner which would guarantee internal temperatures in the range from 15° to 25°C (with humidities below the saturation point) at any injection temperature would provide a major gain in personnel efficiency and health.

Experience with ultra-violet irradiation of circulated air as a method of reducing odor and contagion hazards has been inconclusive. Civilian experience indicates that triethylene glycol may have a more efficient sterilizing action. However desirable sterilization and odor removal may be, it appears that the major need in submarines is for equipment capable of maintaining the thermal balance of the space within the range indicated, and at humidities substantially below saturation.

Suggested Researches

Additional investigations of a basic as well as an applied nature remain to be carried out in order that better planning may be done to improve submarine habitability in respect to temperature and humidity. Basic investigations should include:

1. Researches upon the analysis of the thermal environment. This should include the development of new instruments for the measurement of thermal radiation in the visible and near and far infrared portions of the spectrum. It should be the aim to arrive at an evaluation of the operative temperature of the environment which is the recognized measure of thermal stress.

2. Studies of the sensations elicited in the skin of man by these radiations should be carried out to afford a possible quantitative measure of thermal strain.

3. The effects of sustained thermal stress and elevated body temperature upon sleep and efficiency in mental and physical work should be studied. Some work has been accomplished along these lines, but none of it has been carried out under conditions approximating those to be encountered in the submarine.

Applied researches should include:

1. A project for removal of the cork lagging from a portion of submarine hull to test the effectiveness of this surface in reducing humidity in an experimental compartment. These studies should be carried out with forced air circulation and natural convection, and attention should be directed toward the transmission of sound from the compartment into the sea. The purpose of this study would be to provide a silent, effective means of reducing humidity in the living spaces of the submarine.

2. A job analysis should be made of the cooking aboard a submarine, in order to reduce as far as is practicable the release of

water vapor into the submarine during preparation and eating of meals.

3. A thorough study should be made of the effects of sustained thermal stress and elevated body temperature upon the morale and the health of the crew and the crew's effectiveness in simulated battle practice.

Summary

The primary facts of human heat production have been surveyed in relation to the basic physical laws that govern the thermal interchange with the environment. The physiological mechanisms that achieve temperature regulation of the human body have been described in relation to the three principal phases of this adjustment. Data to illustrate the variation of physiological stress and subjective comfort in different regions of temperature exposure and with different levels of heat production have been presented. In these analyses, the basic equations of physiological heat exchange have been used to illustrate the methods employed in analyzing the physiological adjustments and in assessing the combined thermal effect of environment. Utilitarian standards relating to comfort and the thermal efficiency of clothing have been discussed.

References

1. Adolph, E. F., *et al. Physiology of man in the desert.* New York: Interscience Press, 1947.
2. *American Society of Heating and Ventilating Engineers Guide.* New York: Amer. Soc. Heat. & Vent. Engineers, 1946, p. 60., 1945, p. 86.
3. Bazett, H. C. *Temperature, its measurement and control in science and industry.* New York: Reinhold Publ. Co., 1941. Chapter 6, 489-501.
4. Bedford, T., & Warner, C. G. The globe thermometer in studies of heating and ventilating. *J. Hyg.*, 1935, **34**, 458-473.
5. Benedict, F. G., & Cathcart, E. P. *Muscular work.* Carnegie Institute of Washington, Publ. 107, 1913.
6. Blum, H. F. The solar heat load: its relationship to total heat load and its relative importance in design of clothing. *J. Clin. Invest.*, 1945, **24**, 712-721.
7. British Dept. Comm. on Humidity and Ventilation in Cotton Weaving Sheds. *Minutes of evidence and appendices.* Great Britain, 1909. Home Office.
8. Burton, A. C. Clothing and heat exchanges. *Fed. Proc. Amer. Soc. exp. Biol.*, 1946, **5**, 344.
9. Büttner, K. Bedentung und Messung der Oberflächenfenchte fur Transpirationanalyse. *Biogischen Zentralblatt*, 1935, **55**, 356.
10. Büttner, K. Publication of the Prussian Meteorological Institute; 1934, **10**, 404.
11. Dill, D. B. *Life, heat and altitude.* Cambridge, Mass: Harvard Univ. Press, 1938.
12. Dufton, A. F. The effective temperature of a warmed room. *Phil. Mag.* 1930, **9**, 858.
13. DuBois, E. F. The mechanism of heat loss and temperature regulation. *Lane Medical Lectures.* Stanford U. Press, 1937.
14. DuBois, E. F. *Basal metabolisim in health and disease.* Philadelphia: Lea and Febiger, 1936.
15. Gagge, A. P. A new physiological variable associated with sensible and insensible perspiration. *Amer. J. Physiol.*, 1937, **120**, 277-286.
16. Gagge, A. P., Burton, A. C., & Bazett, H. C. A practical system of units for the description of the heat exchange of man with his environment. *Science*, 1941, **94**, 428-430.
17. Gagge, A. P., Winslow, C.-E. A., & Herrington, L. P. The influence of clothing on the physiological reactions of the human body to varying environmental temperatures. *Amer. J. Physiol.*, 1938, **124**, 31-50.
18. Hardy, J. D., & DuBois, E. F. Regulation of heat loss from the human body. *Proc. nat. Acad. Sci.*, 1937, **23**, 624-631.
19. Hardy, J. D., & DuBois, E. F. The technic of measuring radiation and convection. *J. Nutrition*, 1938, **15**, 461-475.
20. Hardy, J. D., & DuBois, E. F. Basal metabolism, radiation, convection and vaporization at temperatures of 22° to 35°C. *J. Nutrition*, 1938, **15**, 477-497.
21. Hardy, J. D., & Soderstrom, G. F. Heat loss from the nude body and peripheral blood flow at temperatures of 22° to 35°C. *J. Nutrition*, 1938, **16**, 477-497.
22. Hardy, J. D., Milhorat, A. T., & DuBois, E. F. The effect of forced air currents and clothing on radiation and convection. *J. Nutrition*, 1938, **15**, 583-595.
23. Hardy, J. D., & DuBois, E. F. *Temperature,*

its measurement and control in science and industry. New York: Rienhold Publ. Co., 1941, Chapter 6, 537–543.

24. Hardy, J. D. *Human temperature regulation.* Symposium, edit. L. H. Newburgh. Philadelphia, W. B. Saunders Co. (In press)
25. Henderson, Y., & Haggard, H. W. The maximum power and its fuel. *Amer. J. Physiol.*, 1925, **72**, 264–282.
26. Herrington, L. P., Winslow, C.-E. A., & Gagge, A. P. The relative influence of radiation and convection upon vasomotor temperature regulation. *Amer. J. Physiol.*, 1937, **120**, 133–143.
27. Herrington, L. P. Basic procedures in the calculation of the heat exchange of the clothed human body. *Yale J. Biol. Med.*, 1947, **19**, 735–755.
28. Marbarger, J. P., & Taylor, C. L. The effect of air movement on human response to heat and humidity. *Fed. Proc. Amer. Soc. exp. Biol.*, 1946, **5**, 70–71.
29. Murlin, J. R. Skin temperature, its measurement and significance for energy metabolism. *Ergeb. der Physiol.*, 1939, **42**, 153–227.
30. McConnell, W. J., Houghten, F. C., & Yaglou, C. P. Air motion-high temperatures and various humidities-reaction on human beings. *Trans. Amer. Soc. Heat. & Vent. Engineers*, 1924, **30**, 167–192.
31. Nielsen, M. Regulation der Körpertemperatur bei Muskelarbeit. *Skan. Arch. Physiol.*, 1938, **79**, 193–230.
32. Oppel, T. W., & Hardy, J. D. Studies in temperature sensation. I. A comparison of the sensation produced by infra-red and visible radiation. *J. clin. Invest.*, 1937, **16**, 517–524.
33. Plummer, J. H. Publication of the Climatology and Environmental Protection Branch. Office of the Quartermaster General. Aug. 25, 1944.
34. Ranson, S. W. The hypothalamus as a thermostat regulating body temperature. *Psychosom. Med.*, 1939, **1**, 486–495.
35. Report by the Housing Commission. *Quarterly Bulletin of the Health Organization of the League of Nations*, 1937, **6**, 505.
36. Sherman, H. C. *Chemistry of food and nutrition.* VII ed. New York: Macmillan, 1946.
37. Soderstrom, G. F., & DuBois, E. F. The water elimination through the skin and respiratory passages in health and disease. *Arch. intern. Med.*, 1917, **19**, 931–957.
38. Taylor, C. L., & Marbarger, J. P. Some effects of extreme heat and humidity on men. *Fed. Proc. Amer. Soc. exp. Biol.*, 1946, **5**, 104.
39. Winslow, C.-E. A., Gagge, A. P., Greenburg, L., Moriyama, I. M., & Rodee, E. J. The calibration of the thermo-integrator. *Amer. J. Hyg.*, 1935, **22**, 137–156.
40. Winslow, C.-E. A., Herrington, L. P., & Gagge, A. P. A new method of partitional calorimetry. *Amer. J. Physiol.*, 1936, **116**, 641–655.
41. Winslow, C.-E. A., Herrington, L. P., & Gagge, A. P. The determination of radiation and convection exchanges by partitional calorimetry. *Amer. J. Physiol.*, 1936, **116**, 669–684.
42. Winslow, C.-E. A., Herrington, L. P., & Gagge, A. P. Physiological reactions of the human body to varying environmental temperatures. *Amer. J. Physiol.*, 1937, **120**, 1–21.
43. Winslow, C.-E. A., Herrington, L. P., & Gagge, A. P. Physiological reactions of the human body to various atmospheric humidities. *Amer. J. Physiol.*, 1937, **120**, 288–299.
44. Winslow, C.-E. A., Gagge, A. P., & Herrington, L. P. The influence of air movement upon heat losses from the clothed human body. *Amer. J. Physiol.*, 1939, **127**, 505–518.
45. Winslow, C.-E. A., & Gagge, A. P. Influence of physical work on physiological reactions to the thermal environment. *Amer. J. Physiol.*, 1941, **134**, 664–681.
46. *Ventilation report of the New York State Commission on Ventilation.* New York: E. P. Dutton, 1923.
47. Yaglou, C. P. The significance of clothing in air conditioning standards. Address. Annual Meeting, Amer. Hyg. Assn., New York, Hotel Pennsylvania, June 5, 1940.

Chapter 14

DIET

JOSEF BROŽEK and OLAF MICKELSEN

University of Minnesota

Introduction

It is well known that a quantitative as well as a qualitative deficiency of food results in both a physical and a psychological deterioration. However, it is easy to overdramatize the physiological importance of specific foodstuffs. The exaggerated advertisements indicating the benefits to be derived from vitamin pills, bran cereals, and other specific foods are a prostitution of nutritional research and do much to discredit it as a fledgling science.

The overwhelming increase in the number of scientific reports and the morass of nontechnical articles on food might tempt the commanding officer of a submarine to follow the philosophy expressed by the poet who wrote (**35**):

> Methusalah ate what he found on his plate
> And never, as people do now,
> Did he note the amount of the calorie count—
> He ate it because it was chow.
> He wasn't disturbed as at dinner he sat
> Destroying a roast or a pie,
> To think it was lacking in granular fat
> Or a couple of vitamins shy.
> He cheerfully chewed every species of food,
> Untroubled by worries and fears
> Lest his health might be hurt by some fancy dessert,
> And he lived over 900 years!

As far as one can judge from the few available reports on the foods supplied to submarines, there would at first glance appear to be very little room for nutritional improvement. One must realize, however, that many things may happen both to the food before it is eaten and to the sailor who continues to live on this food. For example, unappetizing meals resulting from poor planning of the menu or unsatisfactory ways of preparing the meals may lead to a marked deterioration of morale and, possibly, physiological disturbances as well. There are reports of submarine crews that had to subsist for the latter part of a patrol on very restricted diets; in some cases the men received very little besides beans for the last two weeks of a cruise.

It is a common experience that a heavy meal reduces both alertness and capacity to concentrate. No scientific research has been done on this subject. The submarine, equipped with many instruments, requires a high degree of mental efficiency at all times on the part of certain crew members. Since there are few things for the submariner to really look forward to besides his meals, he is likely to indulge in excess eating. This habit, in addition to the immediate disturbing effects noted above, might affect permanently the individual's dietary practices. Over-eating may mean very little in the twenties, but the resulting overweight present during the next two decades might mean the difference between health or invalidism due to cardiovascular disturbances.

There is at present no nutritional research related specifically to submariners. The problems of submarine life must be projected onto the general background of available nutritional knowledge. On this basis, an attempt will be made to point out those problems, both technological and biological, that demand further attention and are of particular interest from the point of view of undersea warfare.

At the same time we wish to emphasize that improvements in submarine nutrition are intimately linked to the progress of the general science of nutrition. Naval support for research on the broad, fundamental questions of human nutrition, paralleling work on the immediate and narrow problems, will in the long run pay dividends.

In a national emergency, there is an urgent demand for information on the relationship between fitness and diet, but there is little time or facility for a thorough study of the problem. In peace, there is a tendency to relapse into routine activities and limit research to a minimum. In some areas of the study of man, the potential contribution of research is ignored entirely. This seems largely to be the case as far as submarine nutrition is concerned.

Trends in Nutritional Research

In its relatively short history, nutritional research has at various times changed its course rather abruptly and followed narrow channels which were almost completely abandoned in the next phase. Most of the early work in this field was devoted to a study of food as the fuel supplying the body's demands for energy. Toward the end of the last century various types of calorimeters were constructed by means of which laws of energy transformation were studied in living organisms. This work involved essentially the adaptation to nutritional studies of various pieces of equipment and experimental techniques which were already being used in other scientific fields.

At about the same time that these studies were going on, Kjeldahl developed a relatively simple means for the determination of nitrogen in organic substances. Nitrogen is the element which characterizes proteins. The Kjeldahl procedure thus opened the way for extensive studies on this important dietary constituent.

Studies were also made on the more common minerals such as iron, sodium, potassium, calcium, and phosphorus, since the chemical and physical methods of analysis available at that time could easily be used for such purposes. The study of those minerals that are required by the body in smaller amounts had to await the development of sensitive chemical tests and further progress in the general field of nutrition.

The vitamin field showed its greatest advances when it became possible to use purified foodstuffs that were adequate in all respects but that of vitamins. To rations which were practically devoid of any vitamins were added the purified fractions of those foods which were particularly potent in the substance under investigation. By repeated physical and chemical fractionation, many vitamins were finally isolated in pure form. As each new vitamin became available, the recognition of additional ones was that much easier.

Some of the microbiological techniques developed for vitamin assays have proved very amenable to the study of individual amino acids. The development of this technique, together with a number of other factors, has resulted in a marked resurgence of interest in the protein field. The availability of various isotopes will complement and modify the stream of nutritional research in the future.

As the various nutritional factors have been uncovered, their requirements by human beings became a matter of concern both from a qualitative and a quantitative standpoint. It has only been within very recent times that well-controlled studies on human subjects were possible. Previous to that, what little work was done in this area was limited to field surveys. Under such conditions, the criteria used in assessing the nutritional status of the subjects are of considerable importance. With time, the criteria originally set up as bench marks for the various deficiency states have undergone considerable revision.

Corneal vascularization, cheilosis, rhagades, and related symptoms have been considered as pathognomic of a riboflavin deficiency. Now it is recognized that other factors may be responsible for the development of these signs (**40, 58, 59**). Criteria of other aspects of the nutritional status are also being subjected to a more searching analysis and reevaluation. Even such an apparently well-established base as the height-weight relationship is being questioned as a reliable reference point. There are indications that, if methods could be developed for the determination of the amount of fat in the body, then a better basis could be provided for evaluating obesity. Valid information on the body fat content would be especially important for individuals who lead sedentary lives. Under such circumstances, the gross weight might be within the normal range according to the height-weight tables, yet the individual might have undergone a profound physiological change. Much of his original muscle tissue might have been replaced by fat.

The implications of such a change from a nutritional standpoint is becoming clearer with the accumulation of data provided by animal experiments and clinical research. The animal studies concerned with the influence of the rate of growth and the adult weight agree in the following conclusions: (1) Those animals that grow at a slow rate because of a caloric restriction, live longer than the controls fed *ad libitum*. (2) The slower-growing animals show a lower incidence of tumor growths. (3) The caloric restriction necessary to produce these changes is very drastic (**12, 43**). Insurance statistics indicate that overweight in man is associated with higher morbidity and mortality.

The sedentary individual is not only inclined to an accumulation of adipose tissue, but exhibits a concomitant decrease in physical "fitness." His endurance for physical exertion is reduced. There are many suggestions that with this decrease in physical "fitness" there is a corresponding decrease in the ability to withstand various physiological stresses such as infectious disease, physical trauma, etc. (**16**). It must be recognized that at present this thesis is not supported by adequate evidence; it is one area in which rigorous research is urgently needed. The problems are important for the submariner, since his is a relatively sedentary existence, and, unless special measures are taken, some of the debilitating changes may occur prematurely in him.

Differences in Nutritional Requirements

The need for additional nutritional research becomes more apparent as the differences between man and animals are more clearly recognized. Some of these differences will be mentioned in order to emphasize the caution needed in applying the results of animal experiments in human beings. Marked differences such as those that exist between man and the herbivores are readily recognized and accepted. For example, it is well known that the bacterial flora of the cow's digestive tract will synthesize practically all of the vitamins of the B complex. This makes cattle independent of a dietary source of these vitamins.

There are marked differences in the requirements for vitamins and amino acids among the various species of animals. When the differences between species are small, there is a pronounced tendency to translate the findings as indicating *identity* in requirements and physiological responses. The very existence of spontaneous food allergies in some individuals and their absence in animals should be a sufficient deterrent to those who consider animal experiments an adequate substitute for studies on normal human beings. Finally, it should be emphasized that almost all nutritional work is done with animals that have been highly inbred so that the genetic variability was practically eliminated. Such work is necessary for a study of the influence of separate factors. However, any group of men is very heterogenous, even when every effort is made to choose only those who most closely fit a narrow pattern of age, race, mode of life, etc.; under such conditions "general laws" become only gross approximations applicable to the "average man," at best.

Psychologically, the animal studies suffer from serious limitations. One might reasonably ask, "How much information about the influence of diet on special senses, motor and intellective performance and personal and social adjustment can be secured from animal experiments?" Even when such experiments can be carried out under ideal conditions, how applicable are such results to man? It is highly doubtful whether work with animals will ever answer such a question as, "Why does a man refuse sausage, giving as his excuse that it looks like horse meat?" Man's appetite, food habits, and preferences are conditioned by his individual experiences and cultural patterns. Animal experiments can shed little light on these factors.

The psychological alterations resulting from dietary abnormalities usually appear before any biochemical or physiological changes can be detected. When experimental subjects were fed a diet that was very low in thiamine, they early showed marked changes in personality and behavior (**9, 26, 59, 60**). Among the personnel on a submarine who are exposed to many routine "stresses," any dietary dissatisfaction may have a very adverse affect on morale. It matters little whether the complaints are based on physiological or psychological grounds. Actually, the subjective dissatisfaction with the diet is far more important than the question of nutritional deficiencies.

In addition to the differences between man and animals, and between different animal species, the intra-species variability should be taken into account more seriously than it has been so far. Work with different strains of rats has shown considerable differences in the requirement for such vitamins as riboflavin and thiamine (**39**). Definite differences between normal young men were observed in the urinary thiamine excretions (**47**). Even though all of the subjects received the same vitamin intake, there were some individuals who consistently excreted two to three times as much thiamine as others over the entire eight months of the experiment. Do these differences in vitamin excretion reflect some profound physiological differences that escape detection at the present time?

With reference to the caloric intake, the principle of individual requirement is well recognized, the calories required being proportional to the total body metabolism. In connection with other dietary requirements, such as those for vitamins and minerals, the thinking is in general not so clear. Pett (**49**) pointed out that dietary "requirements" should not be indicated by a single value but should be thought of and represented in tables by the population mean and the "normal" range. The determination of dietary requirements for a given person is, to a certain extent, an individual, clinical-experimental problem.

Methodology of Nutritional Research

Theoretical considerations as well as actual experience in nutritional research demand a multidisciplinary approach to the study of human nutrition (**10**). In experimental work where the social and cultural factors are eliminated by being held "constant," a combination of analytical and descriptive techniques is required to define precisely the nutritional regimen, as well as the biochemical, physiological, and psychological effects resulting therefrom.

Laboratory nutritional research is of necessity abstract and can study only limited aspects of the total role that food plays in man's life. The investigation of such problems as food habits, preferences, and aversions must extend to the home, the school, the industrial canteen, the Navy mess hall, etc. A combination of the techniques of social psychology and cultural anthropology in the study of food with reference to the total personality is imperative. The psychologist is trained to approach behavior quantitatively and to pay attention to interindividual variability of the responses. Due to the fact that his science was rooted in psychophysics and due to the temptation to derive "general psychology" from observations on college students, the psychologist is less sensitive to the cultural factors in behavior. Traditionally, the anthropologist's descriptive and qualitative approach has been oriented toward seeing a particular feature of man's life within the framework of the total culture.

The research on food habits made marked progress during the Second World War as a result of the work of The Committee on Food Habits of the National Research Council (**64**). The Committee realized clearly that problems of human nutrition, including food habits, cut across many sciences, and that their study demands the pooling of a variety of research techniques. The Committee's report recommended that data serving to identify the cultural context within which any particular study is made should be recorded routinely (**63**).

McCay (**46**) indicated a need for a prolonged dietary study among the submariners. This should be undertaken by a team of research workers competent to describe the nutritional, physiological, psychological, and social aspects of feeding and eating within the framework of the overall life aboard submarines. Both the human and the technical aspects should be studied. A submarine offers an ideal situation in which to study food consumption over an extended period of time. Such work should include a record of the food issued to the submarine at the start of the cruise, the food actually eaten by the men, waste, refuse, and the food remaining at the end of the study. An attempt should be made to determine whether any aversions to or desires for various foods developed among the men as the study progressed.

The emotional value of food and eating assumes particular importance under conditions of living in very close quarters, with inadequate recreation. The observations of such aspects as the frequency of meals and snacks should be supplemented by personal interviews directed toward the assessment of the psychological significance of food. As far as we were able to determine, there is no published or classified report on submarine cruises which would provide information on the changes in body weight, the simplest index of the quantitative changes in food intake.

Most of the purely technical details of a major study could be investigated ashore. In connection with his studies on messing aboard an aircraft carrier, McCay (**45**) recommended the use of an experimental galley for research aimed at the improvement of equipment, foods, operations, sanitation, and storing. Such a galley, although located at a shore station and attached to a regular mess hall, would reproduce in design the deck space and equipment actually used on ships. This is essentially the principle of "job miniature situations." Its usefulness was demonstrated in a variety of problems, such as redesigning airplane cockpits (**24**) and visual research (**11**). It is well applicable to studies on the technological aspects of meal preparations aboard a submarine.

The principle of utilizing realistic working conditions applies equally well to the training of cooks. Since the space in the submarine galley is very limited, the final training of the cooks should be under conditions which exactly duplicate in all respects that with which they will later have to contend. The men have to be trained to work in close quarters and to utilize all available space most efficiently. Time will eventually overcome the handicap of such inexperience, but the trial-and-error method is not compatible with the required efficiency of cook and crew.

Calculated and Determined Dietary Composition

Valid nutritional research requires that the dietary composition determined by tabular calculations be supplemented by direct analyses of actual meals served. McCay (**44**) states that the standard tables usually give only the average value for the concentration of each foodstuff; the foods actually used in any meal may differ markedly in composition from this standard. Furthermore, the analytical results on which most tables are based were obtained from unprocessed foods. Usually no corrections are provided for vitamin and mineral losses that occur during storage, preparation, cooking, or holding the food. When such corrections are given they may vary considerably from the actual losses.

An idea of the magnitude of the error resulting from the application of the tabular values to mass feeding may be obtained from investigations carried out by McCay during the war. On the basis of these data, Pritchett (**51**) calculated that the differences between the estimated and analytical values were: for calories, analytical values were from 3 to 45 percent less; for proteins, from 14 percent more to 38 percent less; for fats, from 9 to 60 percent less; and for carbohydrates, from 11 percent more to 41 percent less. On an average, the analytical values were less than the estimated results by 6.3 percent for protein, 33.9 percent for fat, 10.5 percent for carbohydrates, and 20 percent for calories. Somewhat similar discrepancies between the calculated and determined food values have been observed by others (**28, 30**).

Psychological Aspects of Food

Although in current research most attention is focused on the nutritional and physiological functions of various foodstuffs, this should not obscure the older and perhaps more basic aspects of eating behavior. Man has developed certain food tastes and patterns, preferences and aversions. To what extent these have been influenced by nutritional needs, facilities for growing and preserving food, mythological concepts, and pleasurable sensations derived from eating is a debatable question. There is no doubt, however, that the sensual appeal of foods has been of considerable importance. Even today, when nutritional information is so common, most individuals choose their foods more on the basis of preference than on nutritional adequacy.

The significance of the non-nutritional aspects of diet becomes more apparent when other factors in the daily routine of living leave one with little else than his meals as a welcome break in a monotonous existence. Pritchett (**51**) pointed out that, "... Men do eat as a pastime and for enjoyment when on extended cruises as they have little else to look forward to until they hit port again. The psychological and morale factors, therefore, must also be considered in any extended feeding period." This has been emphasized with reference to submarine warfare in an earlier report by Leiter (**37**), who stated that, "Submariners are envied by all other branches of the armed forces because they have always received special attention as regards their foods, clothing, physical condition, etc. Food has always received more than its just share of attention and it is not unusual to find such delicacies as lobster, crab meat, sardines in olive oil, canned turkey or crushed walnut ice cream topping in most submarine galleys. Evidently ... no efforts" have been spared "to get the Submarine Service the very best in foods." In spite of the attention lavished on food, according to this report there were many complaints directed at the meals served during patrols. It is difficult to determine how much of this dissatisfaction rested on esthetic, and how much on nutritional, grounds.

Recognizing the importance of taste and of food preferences in dietary practices, one cannot completely ignore the nutritional aspects. It is obvious that there are extremes in both directions. Although sweets might be very appealing to many individuals, it would not be very long before some nutritional disturbance became manifest if the diet were primarily built around these foods. On the other hand, it would be very difficult to get normal young men to consume for any length of time a nutritionally balanced diet composed of a purified protein such as casein, carbohydrate, and fat, supplemented with all the necessary minerals and vitamins.

The question of food acceptability plays a large role in the overall evaluation of the diet. Up to the present time, little work has been done on this problem. As in other phases of nutritional research, realistic, long-term studies are the only means by which useful, practical information can be obtained. Pritchett (**51**) emphasized this point when he said, "Far too often tests and decisions are made on the basis of just tasting and not of consuming the item in prescribed amounts at the same frequency and for the same length of time as the people who have to subsist on it. It is far easier to prescribe food than to eat it."

One of the complaints that has frequently been heard about all of the services is that, even though they received the best food materials available, the dishes served from the kitchens were often of poor quality. To some extent this was true also of certain items prepared by food processors during the past war. Even though the foods met the specifications, they were not as tasty as the same substance prepared by other manufacturers. For this reason it would be advisable to establish a central purchasing agency through which all food requisitions would be channelled. This agency should be charged with the responsibility of supplementing the presently available specifications with some check on the gustatory appeal of the foods. In this respect, due consideration should be given to the presently available techniques for choosing the persons with superior taste discrimination for such positions (**33, 34**).

Nutritional Requirements of Submarine Personnel

There is no need to be concerned about the adequacy of caloric intake except under emergency conditions. Actually, both among submarine personnel and the American people at large, overeating and overweight may present a much more serious problem than undernutrition. Due to the nature of the work and the confining life it entails, there is a possibility that for some submariners eating may serve as a mechanism of escape from monotony. Over a period of time these men may show a considerable increase in weight.

There are no data on changes in body weight, but there is some evidence of a tendency for the submarines to dispose of larger amounts of food than are required for weight equilibrium. In one study involving over 70 men for about 40 days, the caloric value of the food issued per men per day was about 5,200 Cal.; this caloric value was obtained on the basis of the U. S. Dept. of Agriculture Tables of Food Compensation (**66**). The study did not take into account wastage. There is no way of determining how many calories were actually consumed. At any rate, the caloric value of the food "as issued" is more than twice as high as the estimated caloric expenditure of a comparable submarine crew. The average oxygen consumption over a 24-hour period for submarines operating in temperate zones is equivalent to about 2,400 calories per man per day (**4**). There are limitations in the accuracy of both of these calculations, but such caloric differences are worthy of further investigation.

On the basis of the present evidence, there is no reason to think that the nutritional requirements of the submariner are different from those of other moderately active individuals. At first it might appear that, at least in the case of vitamin D, an extra amount should be provided. Even though the submariner may spend a considerable fraction of the year away from the natural source of ultra violet rays, it is doubtful if any disturbances could result therefrom. The adult requirement for this vitamin has not been adequately established and most of the evidence indicates that normal young men can be adequately maintained on the small amounts that occur in the ordinary mixed diets (**55**).

How about the other vitamins? A number of the vitamins in the B-complex group, together with vitamin C, are easily destroyed both in storage and during the preparation of foods. Even when such a relatively staple food as flour is stored at temperatures approaching 100° F., considerable amounts of the thiamine added for fortification are destroyed. In the preparation of foods, important losses of the vitamins may occur if the food is kept hot for a matter of hours. In both of these cases, proper attention to relatively minor details will go far in reducing the nutritive loss. Of as great significance as the loss of vitamins is the change in the taste of such foods. Although very little work has been done in this field, a statement by Dr. Faith Fenton is suggestive of the possible implications. In one of her Office of Scientific Research and Development reports, she wrote that the dehydrated cabbage with the highest vitamin C content on reconstitution rated highest on the taste tests.

Both clinical reports (**2, 25, 38, 61**) and experimental studies (**6, 32, 54, 60**) indicate that prolonged or very drastic reduction in the intake of the vitamins of the B-complex results in profound alterations in body chemistry, work capacity, and personal adjustment. However, unless the submariner selects a very peculiar diet, and it should be noted that one has to go to great lengths to provide a diet severely deficient in the B vitamins (**1**), there is little danger of developing a B-complex deficiency.

It was originally thought that a moderately low intake of the vitamin B complex for a period as short as two weeks was likely to produce marked deficiency symptoms (**3, 29**). More recent work under well-controlled laboratory conditions has shown that normal young men can maintain a high level of physical activity for more than two weeks with small amounts of thiamine, riboflavin, niacin, and vitamin C in the diet (**13, 31**). Should unforeseen circumstances reduce the crew of a submarine to a very restricted but calorically adequate diet, no physiological disturbances are likely to result during the first two weeks. The psychological repercussions resulting from a monotonous diet are likely to be of far more importance than any physiological effects.

A deficiency of vitamin C eventually produces scurvy, the plague of the sailing-vessel era. The submariner need have no worry on this point. Even if the diet were low in this vitamin during the entire cruise, it is very doubtful whether any deficiency symptoms directly attributable to this vitamin would appear. This has been brought out by the experiment made on a young surgeon who voluntarily lived on a diet devoid of any vitamin C for six months. The subject maintained his full schedule of activities, and only toward the end of the experiment did he show any signs suggestive of a vitamin C deficiency (**14**).

Although night vision may be extremely important to the look-out on a submarine, most of the crew must give constant attention to various lighted instruments and gauges. It is doubtful whether a decreased ability to see at night could result, in the submarine crew, from a deficiency of dietary vitamin A. It might be noted that the question of time needed for the development of deficiency symptoms is at present in a confused state, but there is increasing evidence that normal young men can subsist on a diet very low in this substance for a long time before showing any abnormalities (**41, 67**). Impairment of dark adaptation has been associated with a low intake of vitamin A, but this relationship has been shown to be less close than formerly believed (**17, 27, 56**).

As to the minerals, only the calcium intake might cause some concern because of the small amount of milk in the submariner's diet and the recent increase in the suggested allowance for this mineral (**42, 65**). Part of the anxiety which might develop in this area could be overcome by the formulation of recipes which utilized milk and milk products.

Proteins and fats should pose no nutritional problems peculiar to submarine life. At present there is no evidence to indicate that extra or special proteins are required even for hard physical work (**18, 50**). As long as the caloric requirements are met by a mixed diet, it is very likely that the demands for protein will be taken care of.

Contrary to earlier beliefs, the ingestion of even large amounts of fat will not produce any physiological disturbance such as ketosis. Formerly it was thought that as soon as approximately 50 percent of the calories in a diet were provided by fat, ketosis occurred. This is not the case (**62**). Aside from the fats containing large amounts of the saturated fatty acids (e.g., stearic acid), no investigation has unequivocally shown that one fat is more nutritious than another. There are claims that certain fats such as butter contain relatively large amounts of special acids required by some animals (**7, 8**). On the other hand, investigations have shown no detectable differences in the physiological responses of children when different groups were maintained on butter and oleomargarine (**36**).

Foods for Taste

A wide variety of tastily prepared foods is probably the best insurance against the development of nutritional deficiencies. In addition to that, such a dietetic program pays extra dividends in good morale among the men on board the submarine. The instructors in the cooks' training school should be carefully screened with this qualification in mind. During peacetime, it might be advisable to send some of the instructors on tours of duty at restaurants that are known for their cuisine. Enough emphasis cannot be put on this subject of the preparation of meals (**19**). Even the best quality foods in the hands of an indifferent or incompetent cook may result in dishes that are very mediocre.

The problem of food acceptability received increased attention during the Second World War, and a Food Acceptance Research Laboratory was established at Chicago as a part of the Quartermaster Food and Container Institute for the Armed Forces (**22**). Techniques used in commerical food and beverage testing laboratories (**15**) were subjected to a critical methodological analysis (**21**) and integrated with the methods of consumer surveys (**20**). By refining the psychophysical techniques of the laboratory and devising more reliable quantitative criteria of food acceptability, data might be provided for setting up quality control standards useful in dealing with problems raised by selection of raw materials, processing, packaging, and storing (**23**).

The chemical aspects of gustation and olfaction were discussed in detail by Moncrieff (**48**).

To supplement the usual foods served on submarines, research should be initiated to develop means of prolonging the life of various types of fresh fruits. The various institutes established by the growers of specific fruits are in a position to aid and advise in any such program. There are many indications that fresh fruits are one of the things missed most by the submariners. From the standpoint of morale, it may be advisable to consider the possibility of including these foods on the menu more frequently than is done at present. From a nutritional standpoint, fruits, especially in the fresh condition, can be looked upon primarily as a source of vitamin C. It would be an exaggeration to insist that they be included for this reason alone, since there are many other ways in which this substance can be provided for.

One food which is not being utilized as much as it might be is nuts. Their nutritional value is high, their keeping quality good, and they provide a welcome variety.

Food Serving

Even under ordinary conditions on a submarine, it is impossible for all men to eat at the same time. To provide those men who are unable to eat at the scheduled hours with a warm, nourishing meal may, under certain circumstances, be rather troublesome. Most of the difficulties might be overcome by means of the Meal-pack, similar to that used by air-lines for serving meals while in flight. The submarine cook could prepare food for the men who were absent, and by means of the Meal-pack he could be sure that it would remain in essentially that condition until eaten.

The Meal-pack[1] container consists of an insulated double-walled metal case and a removable pyrex dish from which the meal can be eaten directly. The usefulness of this container has been demonstrated in hospital practice. Unbiased reports indicate that food is kept hot (or cold) for several hours after packing. The containers can be stored in a minimum of space, a point which is of primary importance on a submarine.

Food Supply and Food Storage

In addition to fuel, the supply of food is the factor limiting the radius and duration of a submarine patrol. The need for economic utilization of space is evident. To do this requires the development of space-saving foods and space-saving methods of packaging them.

The ideal in this respect would be dehydrated foods. Certain of these, such as potatoes and cabbage, can be used at present for mass feeding and on rehydration are fully acceptable. However, most of the dehydrated products demand further research before they will be developed to the point of complete acceptability.

Partly dehydrated and frozen foods represent a compromise in this direction. Because of their lower moisture content they take up less space, and, since they have undergone only mild treatment, they show less alteration in taste, color, and odor than do dehydrated foods.

Different meats can be prepared in ways that produce a marked reduction in the space required for storage. Chickens cut up for frying require less space than the intact birds, with no undesirable effect either in storage or preparation. In the same way, boneless beef would affect a great saving of space, but, under the present conditions, its quality is frequently inferior. Additional research should be done on the development of high-grade boneless meats. Under ordinary circumstances, the carcass contains from 25 to 30 percent bone and a relatively small amount of visible fat.

During the past war, reports to the Committee on Food Composition of the National Research Council indicated that there was a large increase in the amount of fat left on the meat by the packers. When the meat was prepared in the camps, much of this fat was discarded by the cooks. If the latter did not trim off enough fat, the service men left it on their plates. The net result of the campaigns to conserve this foodstuff was a large recovery of fat from the garbage cans and grease traps. All of this "recovered" fat could have been removed before the meat was shipped to the service depots, thus saving that much storage space and the time required of the crew for the recovery of the fat.

The magnitude of the wastage of cold storage space by improper trimming of the meat at the slaughter houses can be gauged from the fact that, at a single meal aboard

[1] Manufactured by the Meal-pack Corp. of America, New York, N. Y.

an aircraft carrier, 380 pounds of gristle and fat were discarded from 1,350 pounds of "boiling" beef; the waste represented 28 percent of the raw material (**45**). Campaigns sponsored by government and industry for civilian conservation of fat seems misdirected in the face of such large scale losses.

In the submarine, where space is so critical, the removal of bone and excess fat from the meat would decrease the space this food occupied by approximately 25 percent. This is a conservative figure. Besides saving space, cutting the meat into suitable sizes before it is stored would insure the use of that food in its most desirable form. Leiter's report (**37**) indicated that when sections of meat were too large for the available space in the submarine, the loading crew did an impromptu butchering job with an axe. An unskilled person under such circumstances could do much to decrease the value of even the best grade of meat.

The primary argument against cutting up the sides or quarters of meat before it is actually ready for roasting are: (1) it would shorten the time the meat could be safely stored, and (2) the meat would tend to dry out more rapidly. Recent developments in food technology offer means of overcoming both difficulties. The meat could be cut up under sanitary conditions where the microbial flora were kept to a minimum by means of steri-lamps. When the possibility of contamination is thus reduced and means taken to prevent future ingress of bacteria, the first objection can be disposed of. Subsequent contamination can almost be eliminated by such a wrapper as aluminum foil. In addition to its advantages from the sanitary standpoint, the aluminum foil would be very effective in decreasing the dehydration of the meat. Each cut should be adequately labelled, so that it could be easily loaded, stored, and located on the submarine. This work could be done by a group of men especially trained for the job of trimming, boning, cutting, and wrapping the meat, and the cook's training course could be reduced by that much.

All foods should be packaged in such a way that they occupy a minimum of space. One of the most obvious savings in this respect is the substitution of cans with square sides for the present round ones. Much of the canned goods used by the services is packed in No. 10 cans. Simple calculations show that if square-sided cans were used, eight cans of the same capacity could be put in the present carton that holds only six cans. In addition to this, there would be the added saving of the cartons which would ordinarily be required for the extra cans.

Research in conjunction with the Food Container Institute of the Quartermaster Corps at Chicago should be expanded in this field. Specifications might be readily available for the types of containers and cartons that had been found to stand up longest under difficult transportation conditions.

Central Loading Program

It is readily apparent that the proper storage of even uniform containers requires a certain amount of experience and skill. Where the available space is as limited and irregular in shape as that on a submarine, the skill and experience of the loaders becomes much more critical in determining the amount and the condition of the food toward the end of a cruise. If a small group of men at each of the submarine bases were trained for this particular duty, the food could be loaded economically from the point of view of both time and space. There is a reasonable possibility that a trained crew could get considerably more food into the same space than men with no previous experience in such matters. It is also likely that the material stored by such a crew would be in better condition at the end of a cruise. Similar groups could be attached to submarine tenders and would be responsible for loading all food away from the base.

Leiter's report indicated that for the

Pacific theater approximately $280 worth of food had to be discarded at the end of each patrol. If this food could be saved by proper storage the gain would very likely offset the expense involved.

Well-planned loading operations would provide for the location of the foods in those parts of the submarine where least damage would result from marked temperature variations. Even among canned goods, some foods are less affected by high temperature than others. The same is true also of the dry foods such as flour, oatmeal, cereals, rice, sugar, and coffee.

The present loading practices result, at times, in the submarines' running short of certain items before a cruise is completed. This could be avoided if the food loading were co-ordinated with a central menu planning agency. The foods should be stored in such a way that those items to be used first would be most accessible. The foods to be used next would be next in order. By proper planning, the distribution and arrangement of the foods in the various storage compartments could be adjusted for the different types of submarines.

Central Menu-Planning Agency

At present one member of the submarine complement is assigned the duty of food procurement. According to Leiter, "This vitally important job is held to be an undesirable chore by most officers and they usually waste no time in getting rid of it the moment the opportunity presents itself. Usually the opportunity is 'George', a brand new ensign just out of Submarine School . . . ." The food secured by this officer in all probability bears no relation to the menus planned by the cook, if the latter has done any planning.

Since a large component of the morale of a submarine crew during a patrol depends upon the maintenance of high dietary standards during the entire period away from the base, every precaution should be taken to insure such a condition. One way that both of the above problems could be solved would be by means of a central menu-planning group. Such an organization would be charged with the responsibility of devising nutritionally adequate and appealing menus to cover the entire period of the patrol. The menus obviously would have to be planned only by individuals who had intimate knowledge of, and experience with, the foods acceptable to the men. It is likely, however, that much greater variety and selection would be possible than might appear on first thought. The menus could then be used in planning for the procurement and storage of the food. All of this work would be done well in advance of the next cruise. When the time came to load the submarine, everything would be in order for the proper storage of all items.

A central menu-planning group might serve a number of other purposes. In the first place, it would permit a better integration between the cooking schools attended by the new recruits and the actual working procedures followed aboard the submarines. Secondly, during an emergency when the number of new cooks increases at a rapid rate, the cook himself might be put in a better position if he were armed with a centrally issued menu. Leiter's report on this point shows that new cooks are frequently taunted by the crew to prepare the best foods at the start of the patrol. Should the cook give in to such appeals, he might well run short of certain needed items before the stocks could be replenished. A set of "suggestions" from the central menu-planning agency would give him means for "sticking to his guns" and distributing the various foods over the entire period of the cruise. This would go far to avoid those situations in which the submarine crew was forced to live on nothing much besides beans for weeks at a time. Finally, an integrated system of menu planning and food buying would obviate the possibility of the commissary officer's permitting his personal prejudices or preferences to dictate the foods that were provided for the submarine.

Food Wastage

The waste in the preparation of food will vary somewhat with the size of the crew to be fed. McCay (**44**), in his study of a mess hall feeding an average of 3,700 men, found that about one-seventh of the day's nutrients ("as issued") found their way into the scullery garbage cans. This type of information is important from a practical standpoint. Very little work has been done in this field, and what has been done is difficult to locate. For these reasons McCay's findings will be discussed in some detail.

The amount of solids per day per man in these wastes was 104 grams. This refuse contained the following (in gm.): proteins 10, fat 30, ash 6, carbohydrate 58. The calories amounted to 538. The vitamins lost per man per day in this refuse amounted to 0.4 mg. for thiamine, 0.8 mg. for riboflavin, and 9 mg. for niacin. These data may be interpreted more intelligently when compared to the weight of the day's food as issued. The total amount of solids was 666 gm., which contained 117 gm. of protein, 166 gm. of fat, 34 gm. of ash, and 349 gm. of carbohydrates. The food had a caloric value of 3358 Calories. It had 2.1 mg. of thiamine, 3.9 mg. of riboflavin, and 26 mg. of niacin. In order to secure an indication of the composition of the food served, the above scullery losses should be subtracted from the values for the food as issued to the ship. Since no plate wastage studies were made in the mess hall, these figures on food loss are only minimal.

One of McCay's contributions was to obtain data on what the men were eating and drinking at Ship's Service and at other facilities on the station. The daily food purchased at Ship's Service was estimated as 522 Calories, with 60 percent of the calories provided in the form of candy bars. These snacks compensated for the losses in the caloric value of the meals resulting from the scullery waste.

The food wastage aboard a submarine is likely to differ from that observed at shore stations. McCay's data (**45**) on messing aboard an aircraft carrier probably came closer to conditions in a submarine, even though there are vast differences in the number of men to be fed. In contrast to shore stations, the wastage in the course of preparing the food on the carrier was reported to be small, but no specific figure was given. On the other hand, the waste resulting from discarding food by the crew of the carrier amounted to from 10 to 20 percent of the food served, with bread wastage of the same magnitude.

Emergency Food Rations

Two types of emergency rations should be considered for a submarine crew. The first of these involves foods that can be used in case the refrigerating system breaks down completely. Such a contingency may be a remote possibility, but it should, nevertheless, be considered. For such a time, canned meats, powdered eggs, preserved butter, etc., should be carried. The menus could be devised so that if an emergency did not develop, these foods could be used, to a certain extent, in the meals served during the last part of the cruise.

The other type would be for the time when the men had to abandon ship. Here a great deal of work has been done by both the Navy and the Air Force. Their experience in developing various types of life raft rations might go far in solving this problem.

The literature on survival and emergency rations has been summarized by Scott (**53**) and Berryman (**5**).

Comment and Summary

Shortly after this report was prepared, a paper by Schilling and Duff (**52**) on the food problems of submariners appeared. They studied 1,489 patrol reports and, from the comments of the commanding officers, made various suggestions for improving the habitability of submarines. Their recommendations in many respects coincide with those put forth in the present chapter.

Physiologically, no acute nutritional problems exist as far as the submarine personnel are concerned. The requirements for calories, vitamins, minerals, proteins, and other foodstuffs appear to be the same as those of other moderately active men. Consequently, the improvement in submarine nutrition depends on general progress in the physiological understanding of the role of food in the overall metabolism of the human organism and on technological developments of food growing, preservation and preparation.

It may be useful to point out some of the physiological problems, important for submarine nutrition but not strictly peculiar to submarine conditions, which deserve the attention of research workers.

1. The relationship between food and gas production in the gastrointestinal tract is one problem worthy of investigation.

2. It is common experience that a heavy meal is followed by drowsiness, yet we are aware of no experimental investigations of food intake and its influence on alertness. It was found that taking even a moderate meal by normal young men resulted in significant changes in the electrocardiogram; this indicates that the effects of meals cannot simply be passed by as "subjective." The submariner's eating is irregular and the amounts are apt to be large. In the limited space he has no opportunity to "walk off" a heavy meal.

3. There is a hiatus in our knowledge of the changes occurring, during prolonged patrols, in food preferences, likes, and dislikes, and the over-all position of food in the life of the submarine personnel. When individuals are subject to confining and trying circumstance, some of them may find an escape in eating. If this continues for any length of time, it will eventually lead to the development of obesity. On the other hand, nervousness and anxiety may result in a loss of appetite. This in turn may lead to a loss of weight and a decrease in fitness. As a means of checking these factors, it would appear advisable to make determinations of the weights of the men before and during a long cruise.

4. Body weight is a very gross measure of "nutritional status." In this paper an integrated approach utilizing the techniques of all the relevant sciences of man has been emphasized. Additional work is required in many areas before methods can be developed for the quantitative description of the various facets of the human organism. This is essential before the influence of nutrition can be studied. An example of such a need is a method for the estimation of body fat which requires little time and which does not necessitate the determination of body volume by weighing under water.

5. In some fields the methods are satisfactory but the proper physiological interpretation of the data is not fully established; the uncertainties concerning the significance of urinary vitamin excretion values illustrate this point.

6. Evidence has been presented that animal nutritional studies must be supplemented by experimental work with human subjects. Even in those areas where species differences are not questioned, it is difficult to arrive at valid decisions on the quantitative nutritional requirements of human beings solely on the basis of animal studies. This is true both for the amounts of the different dietary nutrients required and the length of time during which certain minimal levels may be consumed before signs of a deficiency appear. Realistic nutritional research is a costly proposition, but no cheap substitute is available or possible.

Acknowledgment

We wish to acknowledge the helpful suggestions and criticisms offered by: Capt. A. R. Behnke, M.C.U.S. Navy, Bethesda, Md.; Mr. C. S. Pritchett, Subsistence Division, Bureau of Supplies and Accounts, U. S. Navy, Washington, D. C.; Dr. Olive McCay, Animal Nutrition Laboratory, Cornell University, Ithaca, N.Y. We have drawn freely on their ideas; our debt to Mr. Pritchett is particularly heavy. The responsibility for the chapter and its limitations rests entirely upon the authors.

References

1. Anderson, M., Mickelsen, O., & Keys, A. Diets and food items for the experimental production of thiamin deficiency in man. *J. Amer. diet. Ass.*, 1946, **22**, 1–7.
2. Axelrod, A. E., Spies, T. D., & Elvehjem, C. A. The effect of a nicotinic acid deficiency upon the coenzyme I content of the human erythrocyte and muscle. *J. biol. Chem.*, 1941, **138**, 667–676.
3. Barborka, C. J., Foltz, E. E., & Ivy, A. C. Relationship between vitamin B complex intake and work output in trained subjects. *J. Amer. med. Ass.*, 1943, **122**, 717–720.
4. Behnke, A. R. Personal communication. 1948.
5. Berryman, G. H. *Physiologic aspects of survival and emergency rations.* Chicago: Quartermaster Food and Container Inst., 1947.
6. Berryman, G. H., Henderson, C. R., Wheeler, N. C., Cogswell, R. C., Jr., Spinella, J. R., Grundy, W. E., Johnson, H. C., Wood, M. E., Denko, C. W., Friedemann, T. E., Harris, S. C., Ivy, A. C., & Youmans, J. B. Effects in young men consuming restricted quantities of B complex vitamins and proteins, and changes associated with supplementation. *Amer. J. Physiol.*, 1947, **148**, 618–647.
7. Boer, J., Jansen, B. C. P., & Kentie, A. On the growth-promoting factor for rats present in summer butter. *J. Nutr.*, 1947, **33**, 339–358.
8. Boer, J., Jansen, B. C. P., & Kentie, A. & Knol, H. W. The growth-promoting action of vaccenic acid. *J. Nutr.*, 1947, **33**, 369–360.
9. Brozek, J., Guetzkow, H., & Keys, A. A study of personality of normal young men maintained on restricted intakes of vitamins of the B complex. *Psychosomat. Med.*, 1946, **8**, 98–109.
10. Brozek, J., & Keys, A. Interdisciplinary research in human biology with special reference to behavior studies. *Amer. Scientist*, 1945, **33**, 103–111.
11. Brozek, J., Simonson, E., & Keys, A. A work test for quantitative study of visual performance and fatigue. *J. appl. Psychol.*, 1947, **31**, 519–532.
12. Carlson, A. J., & Hoelzel, F. Apparent prolongation of the life span of rats by intermittent fasting. *J. Nutr.*, 1946, **31**, 363–375.
13. Cogswell, R. C., Berryman, G. H., Henderson, C. R., Denko, C. W., & Spinella, J. R. Absence of rapid deterioration in moderately active young men on a restricted intake of B complex vitamins and animal proteins. *Amer. J. Physiol.*, 1946, **147**, 39–48.
14. Crandon, J. H., Lund, C. C., & Dill, D. B. Experimental human scurvy. *New Eng. J. Med.*, 1940, **223**, 353–369.
15. Crocker E. C. *Flavor.* New York: McGraw-Hill, 1945.
16. Cureton, T. K. The unfitness of young men in motor fitness. *J. Amer. med. Ass.*, 1943, **123**, 69–74.
17. Dann, W. J., & Yarbrough, M. E. Dark adaptometer readings of subjects on a diet deficient in vitamin A. *Arch. Ophthalmol.*, 1941, **25**, 833–838.
18. Darling, R. C., Johnson, R. E., Pitts, G. C., Consolazio, F. C., & Robinson, P. F. Effects of variations in dietary protein on the physical well being of men doing manual work. *J. Nutr.*, 1944, **28**, 273–281.
19. Dill, D. B. Food preferences on combat flights. *J. aviat. Med.*, 1947, **18**, 590–593.
20. Dove, W. F. Appetite levels of food consumption: A technique for measuring food in terms of psychological and nutritional values combined. *Human Biol.*, 1943, **15**, 199–220.
21. Dove, W. F. Techniques for measuring changes in flavor acceptability. QMC Manual 17-7, 1945.
22. Dove, W. F. Developing food acceptance research. *Science*, 1946, **103**, 187–190.
23. Dove, W. F. Food acceptability—Its determination and evaluation. *Food Technol.*, 1947, **1**, No. 1, 39–50.
24. DuBois, E. F. The anatomy and physiology of the airplane cockpit. *Aeronaut. Engng. Rev.*, 1945, **4**, No. 4, 1–3.
25. Goldberger, J. Pellagra, its nature and prevention. *U. S. Pub. Hlth. Repts.*, 1927, **42**, 2193–2200.

26. Henderson, P. C. R., Wheeler, N. C., Johnson, H. C., Cogswell, R. C., Berryman, G. H., Ivy, A. C., Friedemann, T. E., & Youmans, J. B. Changes in personality appraisal associated with a restricted intake of B vitamins and protein. *Amer. J. med. Sci.*, 1947, **213**, 488–493.

27. Harris, L. J., & Abbasy, M. A. Dark adaptation test: Its reliability as a test for vitamin A deficiency. *Lancet*, 1939, **2**, 1355–1359.

28. Hummel, F. C., Shepherd, M. L., Goldbraith, H., Williams, H. H., & Macy, I. C. Chemical composition of 22 common foods and comparison of analytical with calculated values of diets. *J. Nutr.*, 1942, **24**, 41–56.

29. Johnson, R. E., Darling, R. C., Forbes, W. H., Brouha, L., Egana, E., & Graybiel, A. The effects of a diet deficient in part of the vitamin B complex upon men doing manual labor. *J. Nutr.*, 1942, **24**, 585–596.

30. Kaser, M. M., Steinkamp, R. C., Robinson, W. D., Patton, E. W., & Youmans, J. B. A comparison of the calculated and determined caloric and vitamin contents of mixed diets. *Amer. J. Hyg.*, 1947, **46**, 297–325.

31. Keys, A., Henschel, A., Taylor, H. L., Mickelsen, O., & Brozek, J. Absence of rapid deterioration in men doing hard physical work on a restricted intake of vitamins of the B complex. *J. Nutr.*, 1944, **27**, 485–496.

32. Keys, A., Henschel, A., Taylor, H. L., Mickelsen, O., & Brozek, J. Experimental studies on man with a restricted intake of the B vitamins. *Amer. J. Physiol.*, 1945, **144**, 5–42.

33. King, F. B. Obtaining a panel for judging flavor in foods. *Food Res.*, 1937, **2**, 207–219.

34. Knowles, D., & Johnson, P. E. A study of the sensitiveness of prospective food judges to the primary tastes. *Food Res.*, 1941, **6**, 207–216.

35. La Wall, C. H. The foods of the century. *Hosp. Corps Quart.*, 1944, **17**, No. 6, 1–13.

36. Leichenger, H., Eisenberg, G., & Carlson, A. J. Margarine and the growth of children. *J. Amer. med. Ass.*, 1948, **136**, 388–391.

37. Leiter, Lt. S. S. *A study of some of the problems of nutrition as found on submarines operating in the Pacific Ocean Area.* U. S. S. Aegir (AS 23) Fleet Post Office San Francisco, Calif. Mimeographed report dated May 27, 1945, 42 pp.

38. Lewy, F. H., Spies, T. D., & Aring, C. D. Incidence of neuropathy in pellagra: effect of cocarboxylase on its neurologic signs. *Amer. J. med. Sci.*, 1940, **199**, 840–849.

39. Luecke, R. W., Palmer, L. S., & Kennedy, C. Effects of thiamine and riboflavin deficient diets on rats differing in their efficiency of food utilization. *Arch. Biochem.*, 1944, **5**, 395–400.

40. Machella, T. E. Studies of the B vitamins in the human subject. III. The response of cheilosis to vitamin therapy. *Amer. J. med. Sci.*, 1942, **203**, 114–120.

41. Marrack, J. R. Experimental deficiency of vitamin A in human volunteers. In: Bigwood, E. J., *Enseignements de la Guerre 1939–1945 dans le domaine de la nutrition.* Liège: Desoer, 1947, pp. 261–265.

42. Maynard, L. A. Translating dietary allowances into foods and diets. *J. Amer. diet. Ass.*, 1949, **25**, 18–20.

43. McCay, C. M. Chemical aspects of ageing. In: Cowdry, E. V., *Problems of ageing.* Baltimore: Williams and Wilkins, 1939.

44. McCay, C. M. *Nutrition survey of the mess at Marine Barracks.* Naval Med. Res. Inst., Res. Project X-184, 1944, Rept. No. 2.

45. McCay, C. M. *Observations of the messing aboard an aircraft carrier.* Naval Med. Res. Inst., Res. Project X-184, 1945, Rept. No. 8.

46. McCay, C. M. Personal communication, 1948.

47. Mickelsen, O., Caster, W. O., & Keys, A. The urinary excretion of thiamine as a characteristic of the individual. *Proc. Soc. exp. Biol. Med.*, 1946, **62**, 254–258.

48. Moncrieff, R. W. *The chemical senses.* New York: Wiley, 1946.

49. Pett, L. B. Errors in applying nutrient allowances to dietary surveys or food policies. *Canad. J. pub. Hlth.* 1945, **36**, 69–73.

50. Pitts, G. C., Consolazio, F. C., & Johnson, R. E. Dietary protein and physical fitness in temperate and hot environments. *J. Nutr.*, 1944, **27**, 497–508.

51. Pritchett, C. S. Personal communication, 1948.

52. Schilling, C. W., & Duff, I. F. Analysis of submarine food problems in World War II. *Nav. Med. Bull.*, 1948, **48**, 683–697.

53. Scott, E. M. *Survival rations at sea.* Rept. to Committee Med. Res., O.S.R.D., 1945, Pp. 39.

54. Sebrell, W. H., & Butler, R. E. Riboflavin deficiency in man. *U. S. pub. Hlth. Repts.*, 1938, **53**, 2282–2284.

55. Stearns, G. Vitamin D. *Lancet*, 1943, **63**, 344–348.

56. Steininger, G., & Roberts, L. J. Biophotometer test as an index of nutritional status for vitamin A. *Arch. intern. Med.*, 1939, **64**, 1170–1186.

57. Sydenstricker, V. P., Hall, W. K., Hook, C. W., & Pund, E. R. Amino acids and protein deficiencies as causes of corneal vascularization: a preliminary report. *Science*, 1946, **103**, 194–196.

58. Totter, J. R., & Day, P. L. Cataract and other ocular changes resulting from tryptophan deficiency. *J. Nutr.*, 1942, **24**, 159–166.

59. Williams, R. D., Mason, H. L., Smith, B. F., & Wilder, R. M. Induced thiamine (vitamin B_1) deficiency and the thiamine requirement of man. *Arch. intern. Med.*, 1942, **69**, 721–738.

60. Williams, R. D., Mason, H. L., Wilder, R. M., & Smith, B. F. Observations on induced thiamine (vitamin B_1) deficiency in man. *Arch. intern. Med.*, 1940, **66**, 785–799.

61. Williams, R. R., & Spies, T. D. *Vitamin B_1 (thiamin) and its use in medicine.* New York: Macmillan Co., 1939.

62. Laboratory of physiological hygiene. Unpublished data, 1945.

63. *Manual for the study of food habits. Bull. Nat. Res. Council*, 1945, No. 111, pp. 142.

64. *The problem of changing food habits. Bull. Nat. Res. Council*, 1943, No. 108, pp. 177.

65. *Recommended dietary allowances, revised.* Food and Nutrition Board, *National Res. Counc. Reprint and Circ. Series*, 1948, No. 129.

66. *Tables of food composition.* U. S. Dept. Agric. Bur. of Human Nutrition and Home Economics. Miscel. Public. No. 572, pp. 30.

67. Vitamin A deficiency and the requirements of human adults. *Nature*, 1945, **156**, 11–12.

Chapter 15

THE SLEEP-WAKEFULNESS CYCLE OF SUBMARINE PERSONNEL

NATHANIEL KLEITMAN

University of Chicago

One aspect of the problem of sleep aboard submarines is related to habitability and revolves around the space and comfort of sleeping accommodations. However, there is the much wider conception of sleep as one phase of the sleep-wakefulness cycle, which touches upon emotional stability and adjustment, as well as efficiency in performing the many different tasks assigned to the submarine personnel. The sleep-wakefulness cycle is more closely related to the problems of human engineering than to those of habitability, and it is proposed to devote the major portion of this discussion to the distribution and timing of watches, the hours of meals and recreation, the fixity and variability of daily schedules, and the bearing all these factors have on well-being and efficiency of the individuals concerned.

In most civilian occupations gainfully employed persons work between the hours of 8 a.m. and 6 p.m. Their daily routine, and that of their families, is built around these conventional working hours. Getting-up time is between 6 and 7 a.m.; after-work hours, following the evening meal, are devoted to recreation; and bedtime is 10 p.m. to midnight. Shops, offices, and schools are open during the usual business hours. Even when the breadwinners work at odd hours (in fire and police departments, transportation, theaters, night restaurants, newspapers, two- or three-shift factory operations), other members of their families follow the community pattern of living. By repeatedly observing a definite socially-accepted schedule of hours for work, play, meals, and sleep, one develops a diurnal cycle of sleep and wakefulness, as a result of which it is easier to remain alert during certain hours of the day and harder to keep awake during the fraction of the 24-hour period usually devoted to sleep.

There is nothing physiologically compelling about the 24-hour alternation of sleep and wakefulness. The cycle stems from the astronomical fact of the succession of day and night, to which the social organization of community life is adapted by preference. The individual, however, is born into a society which already possesses a fairly fixed schedule of activities, and is made to "swing into step" with his family and community. Shifts in the timing of the individual's routine occur when he moves into a new time zone, differing by 1 to as many as 12 hours from the old, and in community schedules, through summer and winter times ("daylight saving" and "standard")

A very simple test of the degree of adaptation of the individual to the 24-hour cycle is the fixity of the diurnal body temperature curve. It has been known for a long time that the body temperature of so-called warm-blooded or homoiothermal animals is not really constant, but may show fluctuations of one or more degrees F., without necessarily bringing into play thermostatic physiological adjustments. Among these fluctuations, the most regular one is the diurnal body temperature curve. This curve is not present at birth, but is established and maintained first by the imposition, and later by free acceptance, of the family and community pattern of living. The curve can be shifted, inverted, distorted, shortened, or lengthened, by following a new schedule of activities for a certain number of days or weeks, and it may be completely abolished by disease, when a person is unconscious for a long time (as in encephalitis).

The fixity or repetitiveness of the diurnal body temperature curve has a significance far beyond serving as an index of one's adjustment to the 24-hour astronomical cycle of the earth. Efficiency of performance and the degree of alertness during the customary waking phase of the cycle seen to conform to the rise and fall of the body temperature. Thus, immediately upon awakening in the morning, when one's body temperature is about the same as (or slightly lower or higher than) at the time of going to bed the night before, performance, under laboratory conditions, of any of several mental or physical tasks chosen for testing purposes was found to be as poor as (or slightly better or worse than) at the time of going to bed, whether the criterion of performance was speed or accuracy. During the morning, and sometimes early afternoon, as long as the body temperature is on the up-grade, efficiency of performance increases, reaching either a definite peak or an extensive plateau, very much as does the body temperature curve. Later in the day, sometimes not till the evening, the body temperature curve turns downward, and, with this change in direction, there begins a progressive decline in efficiency of performance. The low at bedtime, as already indicated, is about the same as immediately upon getting up, representing a characteristic "drowsiness level." If one remains awake during the entire night, both temperature and performance fall below their drowsiness levels, reaching their minima between 2 and 4 a.m., when it is hardest to keep awake. Then there begins an upswing which crosses the drowsiness level at the usual getting-up time.

Ship Operation and the Sleep-Wakefulness Cycle

Since operation of a ship under way requires a round-the-clock schedule of activities, the crew works under a timing of watches usually involving a "4 hours on and 8 hours off" routine. With such a schedule, one or another watch section is at a disadvantage with respect to meals and recreation, as well as sleep, since on most surface warships the members of the crew are not allowed to sleep in the daytime. In addition, during general quarters (one hour in the morning and one in the evening) all hands must be at their battle stations. This requires two hours more per day of the other two watch sections than of the 4–8 and 16–20 section which is on watch at that time anyway. To be fair to all, watches are often rotated by the system of "dogging" the watch: the 16 to 20 o'clock watch is split in two, with the new section going on watch at 18 o'clock. From the standpoint of establishing and maintaining a regular diurnal cycle of sleep and wakefulness, the system of dogging the watch at frequent intervals is disastrous. Practically all men retain the shore-type body temperature and efficiency curves, which makes them *sleepy* during night watches. Thus, during these hours of darkness, when sudden attacks are likely to occur and when special alertness is required, the majority of the crew is fast asleep, and the few who are awake are at their least efficiency.

For the purpose of insuring that a portion of the crew is wide-awake and ready for action at any hour of the day and night, the routine that prevails on submarines has several advantages over that of surface ships. There is no reveille and usually no daily general quarters requiring the entire crew to turn out. The men are permitted to sleep during daytime hours, and since practically all work is done under artificial light, the difference between day and night is minimized, or absent. If the commanding officer so desires, the watch may remain fixed for the entire duration of a cruise, or may be rotated at only infrequent intervals. The evening meal hour may be moved from 17:30 to 19:30, thus making it unnecessary to rouse sleeping men for the purpose of relieving the 16 to 20 o'clock watch section for supper. It seems, therefore, that, of all fighting ships, the submarine represents a most favorable combination of factors conducive to the establishment of a *maintained shift* in the sleep-wakefulness cycle. There are, however, circumstances which militate against this development, even on a submarine. The chief among them are the shore-type schedule of meals and recreation, and the inability of the men to have an unbroken stretch of sleep of more than 7 hours' duration. Before suggesting remedies for these unfavorable conditions, it was thought desirable to find out how the personnel of a submarine actually operated, and whether the misgivings concerning the adjustment of the men to a shifted cycle are justified. Unfortunately, only anecdotal information on the subject has been available up to the present. To obtain facts and figures pertaining to the several items of submarine routine under discussion, it was arranged for the writer to participate in a typical cruise of the U.S.S. DOGFISH (SS 350) and facilities were provided for collecting pertinent data.

The DOGFISH is a fleet-type submarine, commissioned in 1946, but recently modernized and converted to "schnorkel" operation. It departed from its home port, the U.S. Naval Submarine Base at New London, Connecticut, for the Bermuda Operating Area on May 5, 1948, returning on May 21. Two week-ends, spent at the U.S. Naval Operating Base at Bermuda, broke the ship's round-the-clock routine of operation. As a result, only nine full 0 to 24 o'clock periods became available for statistical study, but the significance of the mean figures obtained on the duration of sleep is beyond any doubt.

The personnel of the DOGFISH comprised eight officers and 66 enlisted men, among them a rather high proportion of chief petty officers (15). The ages of the crew ranged from 17 to 39 years, with a mean of 25. They had served in the Navy from less than a year to over 18 years, and on submarines as long as 14 years. The mean of naval service of the crew was 7 years; on submarines, over 4 years. Of the entire personnel 43 were married and 31 were single. By a self-appraisal of personality schedule, it was found that 44 were predominately extrovert, 15 intermediate, and 6 predominately introvert (9 preferring not to rate themselves). During the cruise each member of the crew was interviewed by the pharmacist's mate, who had the complete confidence and good will of the men, as to his habits of sleep, meals, recreation, etc., both ashore and afloat, following a definite system of questioning (see Appendix).

There were three watch sections. Section I (2 officers and 16 men) had the 4–8 and 16–20 watches; Section II (2 officers and 18 men), the 8–12 and 20–24 watches; Section III (2 officers and 17 men), the 0–4 and 12–16 watches; and what will be designated in the figures and table as Section IV (2 officers and 15 men) was not required to stand watches, but either followed a frankly shore-type schedule of activities (cooks and messmen) or were on call at all hours. Thus, for purposes of comparison, the entire personnel was divided into four almost equal groups. The hours of watches were fixed for the entire cruise (except for the officers), and the meal hours were made to coincide with watch changes: breakfast at 7:30–8 dinner at 11:30–12, and supper at 19:30–20. The late hour of supper made it unnecessary to awaken Section II men in order to relieve members of Section I for the 17:30–18 supper, as is done on some submarines. But light snacks were eaten prior to 16 o'clock by a considerable number of men of Section I, and also by members of Section III just before midnight.

To make the taking of records simple and expeditious, a large sheet containing the names of the entire personnel was provided for each day's observations. In 24 boxes opposite each name, the individual's activities were recorded by a code system which comprised sleep, lounging, duty, regular meals, snacks, drinking of coffee or tea, and such recreational activities as reading, writing, and playing of games. The writer was assisted in this hourly survey by the pharma-

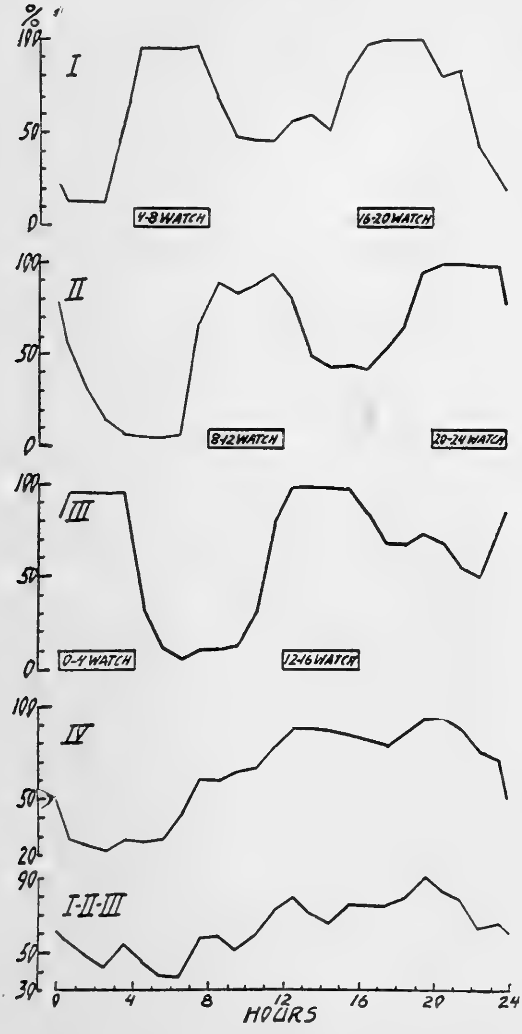

Fig. 1. Parts I–IV show mean percentages of officers and men in each of the four sections awake during every hour. Data are based on nine full 24-hour periods. Bottom curve is a composite of curves I, II, and III.

cist's mate and the yeoman, each of the three operating with one section, on a "4 hours on and 8 hours off" schedule. The hourly note-taking required about one-half hour of the observer's time, which allowed short intervals of rest between rounds. Unless on watch or asleep, each man was questioned as to his activities during the preceding hour and his intentions for the immediate future. Nine individuals volunteered to take their mouth temperature at 4-hourly intervals during the cruise. This permitted the correlation of the sleep-wakefulness cycle with the diurnal body temperature curve.

Routine of Living on U.S.S. Dogfish under Way

Sleep and Wakefulness

As the men were allowed to sleep when off duty, there was not a single hour during which someone was not asleep. However, the distribution of waking hours, plotted in Fig. 1, is definitely skewed, even if the plotting be based only on the three sections who stood watches. In Section I (watches 4–8 and 16–20) between 0 and 3 o'clock only 13 percent of the men were awake, while in the corresponding 12 to 15 o'clock period over 50 percent were up. The same was true of Section II (watches 8–12 and 20–24), were the percentages of men awake between 4 and 7 and between 16 and 19 o'clock were, respectively, 6 and 53. Section III (watches 0–4 and 12–16) largely slept through the breakfast hour, but would get up in time for dinner. Comparing the two-hour periods between 8 and 10 and 20 and 22 o'clock, only 10 percent were awake during the former, while fully 61 percent stayed up after supper. Section IV, which did not stand watches, shows a single wave of wakefulness during the 24-hour period, with a low during the conventional sleeping hours at night. The composite curve of the wakefulness of the three watch sections, which should theoretically give a horizontal line at the 62 percent level (as their average wakefulness comprised 14.9 hours per day) resembles, instead, the curve for Section IV. The diurnal range of wakefulness was naturally greater for Section IV (93 to 21 percent) than for the other sections combined (90 to 36 percent), but the maxima were close and both occurred between 19 and 20 o'clock. In other words, over 90 percent of the entire personnel were up during the supper hours, while less than 35 percent were up between 2 and 6 o'clock.

The diurnal distribution of the sleep periods for the different sections is shown in Table I. Very few men in the watch sections could satisfy their sleep requirements

TABLE I

Distribution of Sleep Periods, Estimates and Actual Mean Durations of Sleep, Self-Ratings, and Preferences

	I 4–8 16–20	II 8–12 20–24	III 12–16 0–4	IV no watches	All sections
Percentage of Distribution of Men/Sleep/Days on Submarine					
Single long sleep periods per 24 hours	16	7	17	39	19
Two equally divided sleep periods per 24 hours	23	23	9	10	17
Two unequally divided sleep periods per 24 hours	47	63	57	37	51
Three or more sleep periods per 24 hours	14	7	17	14	13
Mean Duration of Sleep					
Sleep ashore, estimate in hours	8.10	7.97	8.15	8.13	8.09
Sleep on submarine, estimate in hours	8.53	8.03	8.95	8.44	8.50
Actual mean duration of sleep on submarine	8.56	9.58	9.16	8.96	9.06
Self-Rating					
Highest alertness ashore: in the morning	2	6	6	4	18
in the middle of the day	11	8	8	12	39
in the evening	5	3	6	0	14
late at night	1	1	0	0	3
Fully alert after awakening, on submarine: at once	3	1	0	5	9
in less than 15′	3	8	6	4	21
in 20′ to 45′	10	6	8	6	30
in 1 to 2 hours	3	3	6	2	14
Quality of sleep on submarine: better than ashore	0	0	0	3	3
same as ashore	3	7	4	5	19
poorer than ashore	16	11	16	8	51
Preferences					
Group A, number of men for fixed watch schedules	7	15	13	9	44
Group B, number of men for rotation of watch schedules	12	3	7	8	30
Number of Group A who prefer day duty	0	5	5	2	12
prefer evening duty	0	2	2	1	5
prefer night duty	2	4	2	2	10
are indifferent to hours of duty	5	4	4	4	17
Number of Group B who prefer day duty	4	1	2	3	10
prefer evening duty	2	0	1	2	5
prefer night duty	1	1	0	1	3
are indifferent to hour of duty	5	1	4	2	12
Number of both groups who prefer day duty	4	6	7	5	22
prefer evening duty	2	2	3	3	10
prefer night duty	3	5	2	3	13
are indifferent to hours of duty	10	5	8	6	29

by one long period of sleep, the percentages varying from 7 to 17, but fully two-fifths of the sleep periods in Section IV were of the single continuous type. Dividing the sleep into two equal parts was not a popular procedure, but Sections I and II chose that

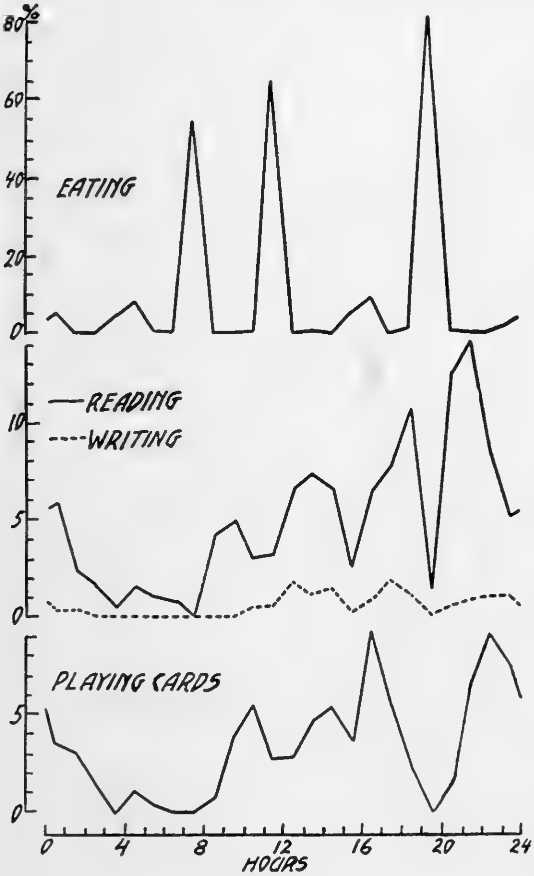

Fig. 2. Mean percentages of the entire personnel of the DOGFISH who were eating, reading, writing, or playing games during every hour. Data are based on ten 24-hour periods. Note the disparity in scale between the upper and the two lower parts.

arrangement over twice as often as Sections III and IV. Taking a nap, in addition to one long sleep period, seems to have been the favorite solution for the three watch sections, and even Section IV indulged in a single additional nap nearly as often as in only one uninterrupted sleep. Finally, there were some cases where three or more periods of sleep occurred in a 24-hour period (13 percent for the entire personnel).

The mean duration of sleep ashore was estimated at about 8 hours per night (Table I). The estimates of the total duration of sleep per 24 hours on the submarine averaged $8\frac{1}{2}$ hours, but data indicate that the mean duration was underestimated by more than half an hour. Peculiarly, the mean durations of sleep were closest for Sections III (9.16 hours) and IV (8.96), and more than one hour apart for Sections I (8.56 hours) and II (9.58 hours). Perhaps the ability of Section II to command the usual night hours of sleep (0–8), and also take an afternoon nap, is responsible for the highest mean duration of sleep for this section. The ranges of the mean total durations of sleep for individuals were about the same in the four sections, actual extremes being, in hours, 7.22–10.44, 7.39–11.28, 6.11–11.28, and 6.78–11.50. Single continuous sleeps of 12–14 hours' duration were not uncommon in Section IV, the only one in which that length of sleep was permissible. Comparing each individual's actual mean duration of sleep with his own estimate of that duration, it was found that two-thirds of the men underestimated and one-third overestimated that figure.

Meals and Recreation

The tendency of wakefluness to be concentrated during the afternoon and evening hours was reflected in the incidence of food intake and recreational activities. Regular meal hours (breakfast before 8, dinner before 12, and supper before 20 o'clock) stand out as spikes in the low curve of food intake (Fig. 2). Whereas fully 80 percent of the personnel showed up for supper, only 65 percent ate dinner, and less than 55 percent got up or remained up for breakfast. Otherwise there was always a spurt in eating at the time of, or shortly after, watch changes, particularly by those who had previously missed the regular meal. Reading and, to a much smaller extent, writing were also engaged in mainly in the afternoon and evening hours. Games, almost exclusively card games, were played very little between 4 and 8 o'clock, and thereafter the curve of their incidence rises to mid-morning, mid-afternoon, and evening maxima. However, reading and card-playing often varied in the opposite sense. Between 16 and 19 o'clock reading was on an increase, while card-playing was falling off. It should be noted (Fig. 2) that the maximum percentage of the men reading at any one hour (21 to 22 o'clock) was less than 15, and of those playing games (16 to 17 and 22 to 23 o'clock) only about 9. Individuals, who did not read, write, or play games while awake, usually lounged in their bunks or engaged in conversation with their shipmates.

Body Temperatures

The nine individuals who volunteered for this chore had their mouth temperatures recorded every four hours. The diurnal curves of four of them, one representative from each section, are shown in Fig. 3. Their hourly variations in wakefulness, superimposed on their temperature curves, show them to be typical of their respective sections in this respect (Fig. 1). It can be seen that the three individuals from the watch sections had double diurnal curves of both sleep-wakefulness and body temperature; that each of the six maxima of temperature occurred toward the end of the watch periods; and that afternoon and evening maxima were generally higher than the night and morning ones. On the other hand, the representative of Section IV (lowermost curve) had a typical minimum body temperature in the middle of his night sleep and the usual temperature plateau between 12 and 20 o'clock. Thus, it is possible, while isolated from external influences on a submarine, both to preserve the shore type 24-hour body temperature curve by following the customary routine of living, and to establish two unequal 12-hour curves, on "4 hours on and 8 hours off" watch schedules.

Self-Rating on Alertness

To determine roughly the degree of adjustment of the personnel of the DOGFISH

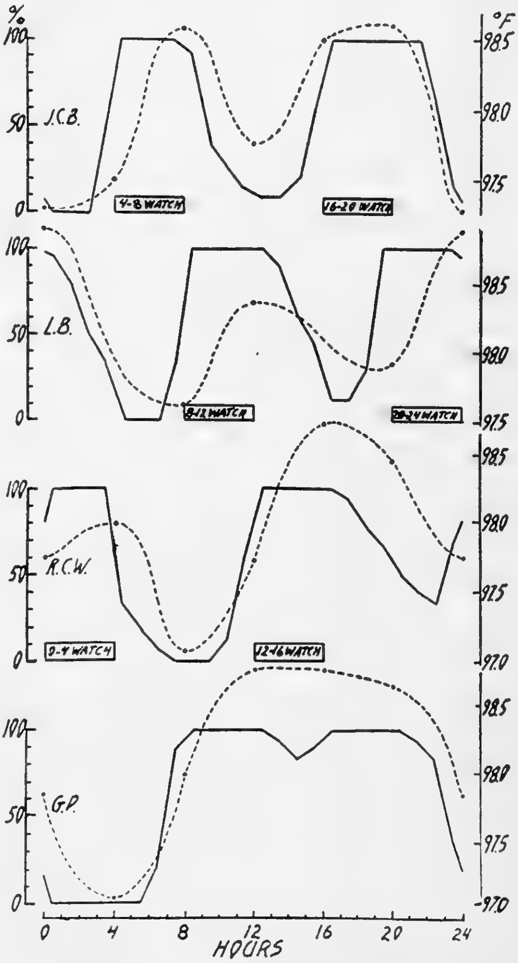

Fig. 3. The percentages of wakefulness for every hour for one man out of each of the four sections of the personnel of the DOGFISH. Superimposed on the wakefulness curves are the mean body temperature curves of the same individuals. Data are based on 9 full 24-hour periods. The mean duration of sleep per 24 hours for each of the four men was as follows (in hours): J. C. B., 9.94, L. B., 8.39, R. C. W., 9.44, and G. P., 7.94.

to the community pattern of wakefulness when ashore, each person was questioned as to the time of the day when his alertness was highest (see Appendix). The self-ratings, given in Table I, reveal few difference among the four sections, 39 in all reporting greatest alertness in the middle of the day, 18 and 14 in the morning and evening, respectively, and only 3 late at night. These figures are about what one could expect to find in a similar-sized group of civilians.

By the nature of their living conditions, civilians are usually not called upon to exercise judgment, requiring full alertness, for one to two hours after awakening in the morning. This time is required for making one's toilet, eating breakfast, and traveling to site of one's job. In the change of watches, however, the men are often aroused 30 minutes or less before their duties begin. They are frequently obliged to skip their meals because of shortness of time. It is interesting to note that, in their self-rating (Table I), only 9 men felt that they were fully alert immediately upon awakening, while another 21 claimed to be so in less than 15 minutes. By contrast, 44 out of 74 required 20 minutes or more to arouse themselves fully, and, of these, 14 needed one to two hours! It is the writer's opinion that, on the whole, the personnel of the DOGFISH overestimated its ability to become fully alert on short notice.

Quality of Sleep on Submarine

The narrow treatment of the sleep problem as a phase of habitability called for an investigation of the sleeping quarters and the type, size, and arrangements of bunks. For the enlisted personnel, there were 17 bunks in Forward Torpedo Room, 18 in Aft Torpedo Room, 27 in the After-Battery Compartment, and 5 in the Chief Petty Officers' Cabin. The bunks were $25\frac{1}{2}$–26 inches wide and 73–75 inches long, altogether too narrow and somewhat short, especially for tall men. The eight officers' bunks were only a trifle wider and longer ($26\frac{1}{2}$–30″ by 78–80″). To make matters worse, some bunks were arranged in three or even four tiers (as in the After-Battery Compartment) which placed them so close to each other vertically that it was impossible for a sleeper to turn over without hitting and scraping against the sagging hammock-type spring of the bunk above. Quite appropriately the men referred to their bunks as "racks." Both the narrowness of the bunks and their vertical crowding discouraged frequent changes in sleeping postures essential for restful sleep. Direct observations of individual sleepers for long periods of time revealed that they tended to "stay put" in the position in which they fell asleep.

Questioning the men brought out practically no complaints concerning their sleeping accommodations. They reported that they fell asleep just as easily on board the DOGFISH as in their shore quarters. However, 51 out of 74 rated the quality of their sleep on the submarine as poorer than that ashore (Table I), and only three men, all of Section IV, said that they could sleep better on the submarine. The absence of complaints can be explained only by the esprit de corps of the submarine service, the general tone of cheerfulness that prevailed on board, and the predominately extrovert type of personality among the crew.

Watch Preferences of the Personnel

Any suggested improvements in the watch schedules must meet with acceptance on the part of the personnel of the submarine. Just as some people persist in eating unbalanced and physiologically inadequate diets because they are used to them, it would be hard, and psychologically unwise, to impose a superior watch arrangement on an unwilling ship's personnel. In answering questions about watch preferences, the men on the DOGFISH expressed themselves, by a majority of three to two, as being in favor of a fixed watch schedule for the duration of a cruise (Table I). However, the four sections did not vote alike. Section IV, which stood no watches, had only an academic interest in the matter of fixity vs. rotation, and they split their vote nine to eight or as close to even as was possible. Sections II and III were definitely for fixed watch schedules, and only Section I voted for rotation. When the men of the latter were questioned on the subject, they gave as their main reason for not wanting to remain on the 4–8 and 16–20 o'clock watches for the entire duration of the cruise that they often slept through dinner, the best meal of the day. By rotation of watches, they would not miss the noon meal for, at least, part of the cruise. The two groups (A and B in Table I) were further queried on their preference for duty at one or another time of the day. Relatively more men preferred day duty in Group B than in Group A, but the opposite was true with respect to night duty. It will also be noted that, of the seven men in Section I who voted for fixed watch schedules, two preferred night duty and five were indifferent to hours of duty. About 40 percent of each group were indifferent to the hours of duty, which is a good basis on which to build physiologically sound watch schedules for the personnel as a whole.

Recommendations for Further Research

The survey of the routine of living on a typical submarine under way revealed an incomplete adjustment of the personnel to the requirement of an even degree of alertness during the 24-hour cycle of day and night and an absence of conditions conducive to such an adjustment. In seeking optimal efficiency around the clock, investigations have to be made of the variation in efficiency under the present schedules of activity, as well as under experimental modifications of current routines. The modifications must be psychologically and socially acceptable and lead to improvement in performance. Both features are susceptible to trial under actual operating conditions of submarines, but individual elements of modified schedules can perhaps be tested more easily under laboratory conditions. The listing of the problems to be attacked under the general heading of achieving greatest efficiency of operation, as well as contentment and health of the personnel, does not necessarily indicate an order of priority in the research projects recommended.

A. Watch Schedules Based on the 24-Hour Cycle

Simplicity of timing operations suggests the preservation of the currently accepted diurnal cycle in the construction of experimental schedules of activities. Fig. 4 represents a schematic arrangement of a routine of watches and meals for each of the three sections into which the *entire* personnel of the submarin should be divided. The time table would b as follows:

Section I

Watches: 8–10, 12–16, 20–22. Meals: 6½–7, 10½–11, 17½–18. Sleep: 22–6

Section II

Watches: 16–20, 22–24, 3–5. Meals: 15½–16, 20½–21, 1½–2. Sleep: 6–14

Section III

Watches: 0–3, 5–8, 10–12. Meals: 22½–23, 3½–4, 8½–9. Sleep: 14–22

Diurnal routine is fixed for duration of cruise, either (1) by seniority, or (2) by individual preference, or (3) by lot.

General features common to all schedules:

1. Three "close" watches during single period of wakefulness: 2, 4, and 2 hours for Section I; 4, 2, and 2 hours for Section II; 3, 3, and 2 hours for Section III.
2. Long stretches of free time: 10 hours for Section I, 11 hours for Section II, 12 hours for Section III; 2 to 4 hours of rest between watches.
3. Meal schedules are arranged so that at least one hour elapses between end of meal and beginning of watch period.
4. Hours of sleep may be optional, falling within the long stretches of free time available under all routines; the hours may also be fixed and, as suggested above, non-overlapping, permitting the exclusion of non-sleepers from sleeping quarters and the use of rotary bedding equipment.
5. At least two hours elapse between awakening and beginning of first watch.

6. All sections can be called to General Quarters between 13 and 15 o'clock without interruption of their sleep periods.

On the proposed schedules the entire eight hours of duty fall within a 12–14 hour period and are subdivided into three unequal, instead of the present two equal, watches. The men are given a full half-hour for every meal and then an additional hour for relaxation prior to standing each watch. The really revolutionary feature of this routine lies in the necessity of preparing and serving nine meals per day. The lower part of Fig. 4 shows how the cooks and messmen, as regular members of the three watch sections, could handle the succession of meals around the clock. Since only one-third of the personnel would have to be fed at each meal, one cook and one messman per section should be sufficient, especially if some cooperative arrangement were worked out among all of them, utilizing individual skills to the fullest extent.

During a trial run under the proposed routine, observations on the distribution of sleep and wakefulness, following the pattern worked out on the DOGFISH, should reveal to what extent the new schedules produce an even degree of alertness throughout the 24 hours of the diurnal cycle. Body temperatures, followed on two men from each section during the cruise, should furnish additional information on that subject. However, neither distribution of wakefulness, nor body temperature data, can be allowed to take the place of actual tests of performance. For these tests, submarine operational procedures, such as diving, firing torpedoes, manning battle stations, etc., should be scored on the basis of speed and accuracy

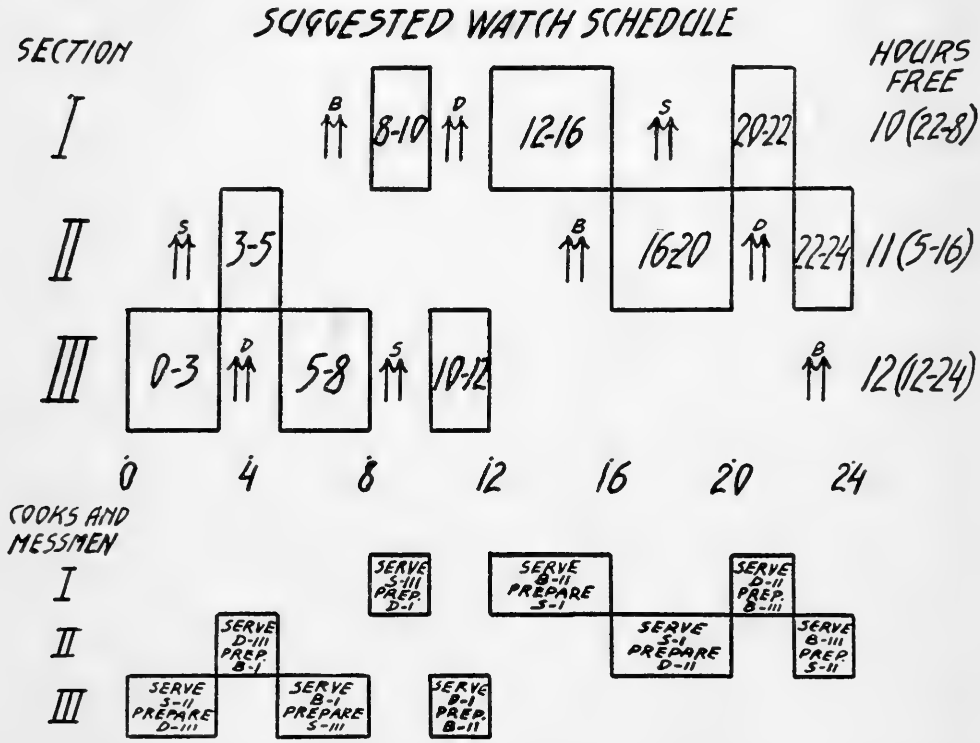

Fig. 4. Experimental scheme of hours of watches, meals, and sleep for the personnel of a submarine. Lower part, a schedule of activities for cooks and messmen who would be required to prepare and serve nine meals to one section at a time, instead of three meals to the entire personnel at once. Each meal (double arrow) is scheduled to last 30 minutes. *B* is for breakfast, *D*, for dinner, and *S*, for supper in each of the three schedules.

under both current and modified watch schedules. The tests should be performed at different hours of the day and night and at various intervals of time following awakening. The new routine has to sell itself by virtue of greater efficiency resulting from its adoption, as well as its cheerful acceptance by the submarine personnel.

B. Watch Schedules Based on a 12-Hour Cycle

The current arrangement of "4 hours on and 8 hours off" lends itself to an investigation into the possible advantages of a frank 12-hour cycle of wakefulness and sleep. The entire personnel would be divided into three sections, operating under the following time table:

Section I

Watches: 0–4, 12–16. Meals: 4½–5, 10½–11, 16½–17, 22½–23. Sleep: 6–10, 18–22.

Section II

Watches: 4–8, 16–20. Meals: 2½–3, 8½-9, 14½–15, 20½–21. Sleep: 10–14, 22–2.

Section III

Watches: 8–12, 20–24. Meals: ½–1, 6½–7, 12½–13, 18½–19. Sleep: 2–6, 14–18.

General features common to all schedules:

1. One 4-hour watch during each waking period of 8 hours; two 4-hour watches per 24 hours: watches of equal duration and spacing under all schedules.
2. Long stretches of free time: eight hours out of each 12, or two 8-hour periods per 24 hours.
3. Four meals per 24 hours, or one every 6 hours, with at least one hour between end of each meal and beginning of watch period.
4. Hours of sleep may be optional, falling equally in each of the two 8-hour stretches of free time between watches or falling predominately within one of the two 8-hour periods. Hours of sleep may also be fixed at 4 hours in the middle of each 8-hour period of wakefulness, permitting the exclusion of of non-sleepers from sleeping quarters and the use of rotary bedding equipment.
5. At least two hours elapse between awakening and beginning of the watch.

In these schedules the timing of the watches would be fixed for the duration of the cruise, but otherwise there are no changes in the now prevailing hours of duty. However, sleep in four or five hour stretches would fall in the middle of the "8 hours off" period, and the culinary adjustment would have to be even more radical than under the first experimental routine suggested: two meals, before and after duty, per 12 hours, for each section, would mean a total of 12 meals prepared and served during every 24 hours. Again, the criterion of superiority of the 12-hour cycle over the present skewed 24-hour one would have to be similar sleep-wakefulness and body temperature curves for all successive 12-hour periods. In addition, greater efficiency would have to be demonstrated by greater speed and accuracy of carrying out typical submarine activities under operational conditions.

C. An Artificial Cycle of Activity

The possibilities of a cycle longer or shorter than 24 hours should be investigated under laboratory conditions. Since submarine operations are likely to be carried out under surface in the future even more than in the past, there is no physiological reason for adhering to the 24-hour alternation of day and night in scheduling the time table of duty, meals, recreation, and sleep. Artificial cycles *can* be established, and the greater temperature swings that are likely to result from cycles longer than 24 hours might lead to greater efficiency during the increased period of wakefulness. The men would still be on duty only one-third of the time, but on a 30-hour cycle would put in altogether 10 hours of duty during 20 hours of wakefulness, with 10 hours given over to sleep. While being tried out under laboratory, rather than submarine-operational, conditions, the same tests could be applied to the advantages of this scheme as have been suggested for those discussed under A and B.

D. Changes in Sleeping Equipment

Although no specific complaints have been voiced by the crew of the DOGFISH, the sleeping bunks were both too narrow and too close to each other. The quality of sleep on wider and longer bunks could be investigated under both laboratory and submarine conditions, using subjective and objective criteria. For the former, reports by men sleeping alternately on bunks of different dimensions would suffice; for the latter it would be necessary to employ devices for recording the frequency and magnitude of movements, as well as the total time spent in motility during sleep. The crowding could be overcome, even without increasing the present allotment of space for sleeping quarters, by redesigning the placement of the bunks. Since, under ordinary conditions, only two men out of three can be asleep, or merely lying down, at one time, an arrangement for having any one of the three bunks out of the way would permit greater vertical space between the other two bunks. A reading lamp at the head of each bunk would be a great boon to those who like to read and are often confined to the only place they can call their own by the crowded and noisy condition of the mess room.

When new submarines are built, serious consideration should be given to enlarging the sleeping quarters of the personnel, both officers and crew, and to changing the design, construction, and arrangement of the sleeping equipment, in line with data obtained by observation and experiment.

Summary

Future studies of the contribution good sleep can make to the well-being and efficiency of the submarine personnel can thus be envisaged as directed along two independent, though complementary, lines: the establishment and maintenance of a regular sleep-wakefulness cycle, in conformity with the prevailing routine of activities, and the achievement of restful and restoring sleep by proper design and arrangement of sleeping equipment.

Acknowledgments

For the survey of other factors related to efficiency of submarine operation the several authors could and did draw on data from a voluminous literature. On the subject of sleep and wakefulness on submarines no published accounts are available, and it became necessary to make a personal survey of existing conditions as a basis for formulating recommendations for future research. The writer, therefore, wishes to express his gratitude to the Committee on Undersea Warfare, the Staff of Commander, Submarine Force, U.S. Atlantic Fleet, and most particularly to all the officers and men of the DOGFISH for their wholehearted cooperation in the collection of the data upon which this survey is based.

References

1. Kleitman, N. *Sleep and wakefulness.* Chicago: Univ. Chicago Press, 1939.
2. Kleitman, N., & Ramsaroop, A. Periodicity in body temperature and heart rate. *Endocrinology*, 1948, **43**, 1.

Appendix

Personal Inventory of Habits of Living—SLEEP, DUTY, MEALS, RECREATION

Name.................... Rank or Rate.......
Serial No......... Length of service: in Navy .... on S/M.......... Married.... Single .... Age....

A. PRIOR TO JOINING NAVY (and when at liberty or on leave, since joining Navy)

SLEEP Usual time of going to bed........ of getting up........

Going to bed later than usual: often.... seldom .... regularly on....

Going to bed earlier than usual: often.... seldom .... regularly on....

Getting up later than usual: often.... seldom .... regularly on....

Getting up earlier than usual: often.... seldom .... regularly on....

Usual state at bed-time: wide-awake.... moderately sleepy.... very sleepy.... varied....

Usually going to sleep: with ease.... with difficulty.... varied.... not noticed....

Usually sleep through night: without waking up .... waking up about .... times
Dreaming: never.... seldom.... often.... every night.... several times per night....
Kind of dreams: pleasant.... unpleasant.... indifferent.... terrifying.... varied....
Manner of awakening: spontaneous (by oneself) .... by alarm clock.... by being called....
Usual quality of sleep: excellent.... good.... fair.... poor.... varied....
Usual state on getting up: well-rested.... tired .... varied.... not noticed....
Highest mental alertness in: morning.... middle of day.... evening.... late at night....
Naps taken: never.... seldom.... often.... daily.... in afternoon.... in evening....
If married, prefer: double bed.... twin beds.... common bedroom.... separate bedrooms....
MEALS Breakfast: large.... small....; lunch: large.... small....; supper: large.... small ....
Kind of beverage with meals: breakfast...... lunch...... supper......
Snacks at odd hours: never.... seldom.... often.... daily.... (number per day....)
Beverages at odd hours: coffee.... tea.... (number of cups per day....) cokes per day ....
Eating before going to bed: never.... seldom.... often.... nightly....; much.... little....
Drinking before going to bed: coffee.... tea....; never.... seldom.... often.... nightly....
RECREATION Active sports.... spectator sports.... movies.... theatre.... reading .... (fiction.... non-fiction.... newspapers .... magazines.... comic books.... other....) dancing.... cards, non-gambling.... other games....

B. SINCE JOINING NAVY, while on active duty on S/M

SLEEP Usual duration of "long" sleep (undressed): .... hours, varying from .... hours to .... hours. Usual duration of "short" sleep and naps: .... hours, usually divided into .... parts. Naps taken: never.... seldom.... often.... daily....
Usual state before long sleep: wide-awake.... moderately sleepy.... very sleepy....
Usually going to sleep: with ease.... with difficulty.... varied.... not noticed....
Usually go through long sleep: without waking up.... waking up .... times
Dreaming during long sleep: never.... seldom .... often.... every time....
Kind of dreams: pleasant.... unpleasant.... indifferent.... terrifying.... varied....
Manner of awakening from long sleep: spontaneous (by oneself).... by being called....
Usual quality of long sleep: excellent.... good .... fair.... poor.... varied....
Long sleep: better than ashore.... about the same as ashore.... poorer than ashore....
Prefer to get up (or be called): as close to beginning of watch as possible.... .... minutes (or .... hours) before beginning of watch
Feel fully alert: at once after awakening....; .... minutes (or hours ....) later
DUTY Prefer: fixed watches for entire cruise.... rotation of watches during cruise.... day watches.... evening watches.... night watches.... indifferent to hours of watches ....
Most efficient: 0000–0400.... 0400–0800.... 0800–1200.... 1200–1600.... 1600–2000.... 2000–2400....
MEALS Prefer to eat at: regular hours.... irregular hours.. regular and snacks....
Number of meals per day: .... (....large, ... small) .... snacks
Number of cups of coffee per day: .... (with meals:.... without meals:....)
Eating before long sleep: never.... seldom.... often.... every time....; much.... little ....
RECREATION Radio.... cards.... other games.... writing.... reading.... (fiction .... non-fiction.... newspapers.... magazines.... comic books.... textbooks.... other........)

C. SELF-APPRAISAL OF PERSONALITY (may be left unanswered)

Always cheerful.... usually cheerful.... variable.... usually gloomy.... always gloomy .... subject to moods.... forward.... shy .... make friends easily.... make friends with difficulty.... like to talk to strangers.... forget wrong quickly.... remember wrong....

PART VI.

HABITABILITY: B. PSYCHOLOGICAL FACTORS

Chapter 16

RELATION OF NOISE TO THE HABITABILITY OF SUBMARINES

F. K. BERRIEN

Colgate University

Research on the effects of noise has been concerned either with the effects directly on auditory functions or on general fatigue and other non-auditory physiological or psychological functions. Conclusions in the first of these fields are fairly clear cut; in the latter, generalizations are hazardous. The applicability of these basic research results to submarines and surface craft is complicated by an apparent variability of moderate magnitude in noise conditions among such ships. Furthermore, few submariners have mentioned noise as an important factor adversely affecting personnel, although it is obviously recognized as an important factor so far as detection and defense are concerned.

The summary of 1471 patrol reports compiled by Duff (**20**) includes only one comment suggesting that noise has adversely affected personnel. One commanding officer noted that engine room men were becoming deaf because of the high noise levels.

On the operational side, a report of factors involved in sonic listening recommends that the sound operator be located in a place as free from airborne noise as possible (**57**). Harris (**33**) examined systematically the effect of masking noise on pitch discriminations of the kind often demanded in sonic listening. Discriminations were poorest when the noise was within five decibels (db) of the signals and improved consistently as the difference between the noise and signal levels became greater.

Although it is beyond the scope of this discussion to deal with the problems of communication, it is important to indicate in passing that noise often masks auditory communication. This difficulty has been highlighted by a report that on one Canadian mine layer, conversation was impossible in the engine room. Sometimes the men in this space missed telegraph signals unless a man was sent from the bridge to alert the engineers (**7**).

The precise noise conditions have been reported for the interiors of only two submarines, although overside measures are apparently made routinely. The earlier of these two reports dealing with USS PERCH indicates that noise levels when submerged approximated 75 to 95 db, depending upon location and speed. However, during the blowing-up operation, when both rapid communication and a high degree of team work are essential, noise levels increased to 102 to 124 db. Under night attack conditions (surfaced) engine room noise varied between 111 and 116 db, although conning and control spaces had noise levels approximating 76 to 90 db (**58**).

Harris and Stover (**34**) not only found the levels somewhat lower on the USS RATON, but also analyzed the noise spectrum. Engine room noise (surfaced) varied between 99 and 108 db. Forward torpedo rooms were recorded as 68 db (RATON) compared with 82 db (PERCH). The authors conclude, "It seems clear that on the RATON there are no noise levels which will be likely to produce detrimental effects on psychomotor performance, or on hearing per se except in the engine room with Diesels operating—and even here the noise level at the most extreme is not such as to produce of necessity any loss of job efficiency. With Diesels operating, the engine room crew will suffer distinct loss of hearing, which will persist to some extent for hours, but, in all but the most susceptible, will undergo complete recovery in time."

Moreover, the noise spectrum peaked at 400 cycles per second (c.p.s.) with very little energy above the 5,000 band. This type of spectrum is not one that would be especially annoying (see later section).

These measurements of noise level are not necessarily typical of all submarines. A report of over-all noise levels of 70 fleet submarines while underway at 40 r.p.m. shows that the USS PERCH is the third quietest (**51**). The RATON was not included in the 70 studied.

Noise Deafness

Goldner (**28**) has distinguished between two types of occupational deafness, acute and chronic, the first being due to blasts and the second resulting from frequent exposures to moderately loud noise. The latter type has a gradual onset and hence is not always noticeable to the individual.

Blast

A great many reports deal with the effect of gun blast and similar high intensity noise of short duration (**31, 36, 38, 40, 41, 54, 56, 62, 64, 65, 87, 88, 91, 98**) to which service personnel have been repeatedly exposed (artillery men, naval gun crews, rifle coaches). In general these reports deal with field or clinical observations of men who have had auditory mechanism complaints. They tend to agree in showing a hearing loss centered in the 2,000 to 4,000 range of frequencies. Some investigators (**77, 88, 98**) have made audiograms before and after exposure to such noises and have thus demonstrated in general (a) a temporary deafness localized between 2,000 and 5,000 c.p.s., and (b) permanent auditory injury in the same range extending fan-wise on the audiogram (increases in spread of frequencies affected and magnitude of loss) with continued exposure and depending upon the duration and intensity of noises. It is to be noted that the range of frequencies thus affected is partially included in the upper range of the secondary speech sounds; and for the most part such injuries, even if permanent, affect only slightly the individual's ability to understand spoken language. However, auditory signals such as those which might be involved in sonar or radio code reception might be adversely affected.

Continuing High Intensities

Insults to the auditory mechanism comparable to those resulting from gun blasts have been discovered in personnel subjected to high intensities of a relatively stable nature over an extended period of time. Such people include flying personnel, tankers, and submariners. The majority of such studies have dealt with flying personnel, although the noise intensities are of approximately the same order as those found in submarines and small surface vessels (**8, 10, 11, 12, 13, 22, 29, 37, 56, 77, 78, 90**).

It is to be noted that the frequencies adversely affected by either blast or high intensity exposures of relatively long duration do not exactly correspond to the predominate frequencies in the exposure noises. Instead, the bulk of reports indicate the major damage centers on 4092 c.p.s. Possible explanations of this finding have been suggested by Campbell (**8**).

The only report discovered which has dealt specifically with submariners covered 1500 otological examinations. It concluded that the major hearing loss due to Diesel engine noise occurs in the first hour of exposure, with further but slight decreases in acuity for two-, three-, and four-hour exposures. Following a one-hour exposure, five hours are required for recovery to normal auditory acuity, while a four-hour exposure often requires as long as 20 hours for recovery. Some evidence was discovered of permanent hearing loss as the individual's exposure in engine rooms was prolonged (**82**).

A laboratory study (**17**) suggested that the higher the frequency of the exposure tone, the greater its effectiveness in producing hearing loss, other things being equal. However, marked individual differences were noted. This laboratory study confirmed the clinical observations noted above in demonstrating that hearing loss is produced most rapidly at the beginning of the exposure period and then more and more slowly as exposure is prolonged. Furthermore, recovery is usually rapid at first and then proceeds more and more slowly, following approximately an exponential curve. However, hearing loss from exposures from 140 db for two minutes is still considerable after two hours, but recovery is nearly complete within 24 hours. This study found no evidence of cumulative permanent effects of 15 or 20 exposures (at intervals of several days at intensity of 130 or 140 db), provided recovery is complete or nearly so after each exposure.

Although he reports only clinical observations, J. A. Weiss (**93**) states "the harmful level of sound intensity is considered to begin between 85 and 100 db for prolonged noises." One of the clinical cases reported in this article is a 37-year-old patient who had a history of duty in a Diesel engine room on a patrol craft for eight hours daily, 24 days per month, for 14 months preceding examination. His hearing loss extended over the entire frequency range and varied from 10 to 55 db.

These laboratory and clinical observations of aviation, surface craft, and submarine personnel are in general agreement with comparable observations made of personnel exposed to manufacturing and construction noises of high intensity. McCoy (**53**) reports audiometric tests of shipyard workers before and after one day's initial assignment as riveters and chippers exposed to rapid impact noises of 110 to 135 db. Average hearing losses were highest in the 2,000 to 5,000 c.p.s. range, varying between 50 and 60 db. Losses in the lower frequencies were 15 to 18 db. Comparable results have been reported in the following references: **6, 27, 66, 72, 73, 74, 89.**

Individual Differences

Throughout the reports on stimulation deafness is an almost universal comment that marked individual differences occur both in regard to susceptibility to hearing loss and to speed of recovery. Thus, Ullman (**90**) found that 30 percent of officers assigned to Flying Fortresses or Liberators were found to have a hearing loss of more than 50 db in at least one frequency, although others exposed an equal period of time showed losses not exceeding 10 db. Senturia (**78**) reports further that his study of 500 pilots before and after exposure to aircraft noise showed, contrary to popular belief, no tendency for those with greatest initial deafness to be unusually sensitive to auditory fatigue. Furthermore, there seemed to be an improvement in hearing in some ears after 70 hours of flight when he compared audiograms at that time with previous measurements made at the end of 10 hours of flight, provided that both sets of measures were taken later than 24 hours following the last exposure. This apparent improvement in hearing has not been reported in other studies, but if substantiated might provide a basis for explaining the apparent adaptation to noise to be discussed in greater detail later.

Non-Auditory Effects

Reports of the effects of noise on vital processes such as blood pressure (**25**), respiration rate (**25, 60**), metabolism (**32, 46, 47**), muscle tension (**18, 19, 26**), and digestive processes (**84**), generally agree that the effects are such that, at least during the initial periods of exposure, physiological "costs" increase. The amount of this increase percentage-wise, its significance, and the speed with which the individual adjusts physiologically to the adverse effects of noise are points of disagreement among the various investigators. In evaluating these studies it must be noted that the noise levels and periods of exposure employed in these studies did not in any instance equal those characteristic of submarine engine rooms when surfaced.

Stevens and others (**85, 86**) concluded that exposure to airplane noise at 115 db for continuous seven-hour periods repeated over 16 experimental days distributed over four weeks had *no effect* on tasks labeled coordinated serial pursuit, serial disjunctive reaction time, card sorting, rotary pursuit, code translation tests, and monocular distance judgments. Measurements on five subjects of metabolic rate, respiration rate, speed of visual accommodation, saccadic movements, bodily sway, hand steadiness, frequency of reversible prospective, and speed of dark adaptation all gave highly variable results from subject to subject and from time to time, apparently unrelated to noise conditions. It was therefore concluded that these tests gave *inconclusive* results on the effects of noise. On the other hand, a coordinated serial reaction time task requiring subjects to direct a beam of light along an opaque pathway by means of airplane controls showed approximately a five percent decrement (statistically significant) in the speed and accuracy with which the task was performed. This result, the investigators suggest, might be an artifact of certain minor auditory cues present during the "quiet" periods but masked by the noise, or of the fact that the subjects set their own pace on this task and worked near the limit of their capacity.

A series of laboratory studies conducted in England (**69**), in which the noise condidions are not completely described, showed that the effects of noise on a simple motor task (placing pegs in a moving peg board) gave results that were difficult to interpret. Two subjects had nine percent fewer misses with noise and two other subjects had 71 percent *more* misses. The authors recognize that they did not adequately control practice effects that might in part account for some of the discrepancies noted. Regular, recurring movements tended to be more disturbed by irregularly presented noise than by noise synchronized with the movements. Mental tasks, such as a number setting test in which the individual operated levers to obtain a particular combination of numbers, showed a small adverse effect. Speed of operations was diminished, although, with continued exposure, the adverse effect of noise was less noticeable. Subjective reports suggested that noise was more distracting in mental rather than on simple motor skills.

A Russian report (**5, seen only in abstract**) suggests that visual efficiency is adversely affected by a 20- to 30-minute exposure to engine noise at 105 to 110 db.

Three field studies of the effect of noise on production (**94, 95, 97**) have agreed in showing a small decrement in output that seemed traceable to the noise factor. However, all of these are open to methodological criticisms and cannot be accepted wholly at their face value (**3**). Weston and Adams (**94**), examining the output of weavers over a 26 weeks' period during which the workers wore ear defenders on alternate weeks, showed an output gain of one percent while wearing ear defenders. Since the output was largely controlled by loom speeds and could not in itself vary greatly, a more revealing statistic was the increase of operators' speed during the five minutes per hour when the weavers were tying knots or making other adjustments. This work, under the control of the operators, increased in speed about 12 percent when wearing ear defenders. A somewhat comparable study (**95**) done by the same authors under similar circumstances showed a $3\frac{1}{2}$ percent advantage in output and a $13\frac{1}{2}$ percent improvement in speed. However, great variability from month to month was shown in speed of operations, which suggested important uncontrolled factors influenced the data. The second study, in contrast to the first, employed two groups of workers, one of which wore ear defenders and the other did not. The equivalence of these two groups initially was never demonstrated.

The third important field study (**50, 97**) was done at the Aetna Life Insurance Company, in which bonus figures, reflecting the efficiency of employees, were used as the indicating data. The installation of a sound-absorbing ceiling in a general office area housing 15 clerks resulted in a 9.2 percent increase in efficiency compared with the year previous during which operations and personnel were identical. Bonus figures during the "quiet" year remained consistently above those during the preceding noisy year. Moreover, when the sound-absorbing ceiling was covered with a hard gypsum board, restoring its reverberating quality, the bonus efficiency immediately dropped to an intermediate value betweeen the quiet and noisy years, but within two months was approximately equal to the level of the quiet year. This suggests that the clerks had readjusted themselves to the higher noise level. At this point the study was terminated due to a reduction in work.

The data of this Aetna study have never been presented in a way which allows for the calculation of statistical tests of reliability of the differences noted. Moreover, the effect of suggestion was not eliminated. It is possible that the clerks may have known the purpose of the study and may unwittingly have exerted greater effort to attain higher levels of production when the acoustic ceiling was installed. The effect of such suggestion has been experimentally demonstrated by Baker (**2**) for short periods of work. Whether such suggestion could operate over an extended period of several months is not known.

Adaptation to Noise

The physiological costs of noise are greatest at the onset of noise and gradually taper off as the noise continues (**18, 35, 48**). The fact of adaptation shows itself not only in such factors as metabolic rate, digestive disturbances and muscle tension, but also in terms of subjective reports of annoyance. The subjects in the English experiment mentioned above (**69**) indicated that they could hardly bear tbe noise in the early experimental sessions, yet at the end they were almost oblivious to it. We have no way of knowing whether the changes subjectively reported are traceable to temporary auditory fatigue in the high frequencies which have been demonstrated to be especially annoying or whether the apparent adjustment has some other basis.

General Fatigue

It has been repeatedly reported on the basis of clinical observations that noise contributes to general fatigue (**3, 4, 16, 32, 46, 84**). This factor has not been systematically investigated, aside from the work on metabolism reported by Laird (**46**) and Harmon (**32**). Laird reported a 19 percent increase in metabolic rate of typists when they were subjected to undampened office noises compared with work in the same space covered with sound-absorbing materials. Harmon's results disagreed with Laird's, but Harmon asked his subjects to engage merely in mental arithmetic in which the muscular component is obviously slight, so that metabolic rate could nor be expected to show much change.

Clinical observations bearing on fatigue take the form of statements such as "standing a watch over a couple of laboring Diesel engines takes more out of a man than when standing watch over a pair of purring turbines" (**21**). A pilot reported, "A person gets much more tired in a noisy plane than he would normally. The physical exertion is not sufficient to warrant this feeling. It must be attributed to the noise which at the end of a long flight leaves my ears ringing and with a definite logy feeling" (**23**). Fifty percent of 65 instructor pilots in PBY's volunteered the comment that noise contributed to unusual fatigue. The existence of such statements drawn from a great variety of sources suggests that there may be some validity to the notion—at least it is a hypothesis that merits more systematic research than it has so far received.

Annoyance

As indicated above, not all noises of a given loudness are equally annoying. Two laboratory experiments (**45, 48,**) in general agree that the high frequencies are more annoying, other things being equal, than middle frequencies (500 to 1000 c.p.s.). This laboratory finding has in a sense been verified by field studies dealing with the degree of acoustical comfort provided by sound-absorbing materials installed on walls and ceilings. Although such installations in areas of high noise levels reduce the overall noise levels only one, two, or three db, people working in these areas generally report a marked decrease in noise annoyance (**4, 63**). Such materials absorb relatively more of the high frequency energy than low frequency energy and consequently produce a noise spectrum in which the high frequencies are markedly reduced.

Interrupted or discontinuous noise has been found to be more annoying than steady noises (**47, 69**). A number of writers have referred to other factors, none of which have been subjected to systematic investigation. Among these are the unexpectedness of noise, the spreading effect, or reverberation, and the degree to which noise is unnecessary or indicates malfunctioning of equipment.

Noise Protection

Efforts to deal constructively with the noise problem may be directed to (a) the suppression of noise at its source by redesign of equipment, (b) the reduction of its transmission through structures, (c) the reduction of reverberation by placing low density materials on walls or ceilings, or (d) the use of protective devices applied to the ear mechanism. It is beyond the scope of this report to deal with the first three of these methods, although the Army Signal Corps recently completed studies (**75**) directed to this end on small Diesel engines.

Investigations of various types of ear protectors has been published (**79, 100**), showing that V-51R type ear wardens are in over-all performance superior. These provide 25 to 30 db acoustic insulation at low frequencies and 40 db or more at the high end of the audio-frequency range. Furthermore, speech communication in the presence of high level continuous noise is not importantly hindered and may be definitely improved by the use of these wardens. Service reports from field tests indicate protection thus afforded against gun blast is "adequate."

Comparative acoustic insulation tests were performed with other types of ear defenders and showed that the V-51R gave generally greater acoustic insulation with least distortion over the audio-frequency range. The author concluded, "A signal that is audible to the open ear in a noise higher than 75 db is equally audible or more audible when ear wardens (V-51R) are worn. Noise levels in submarine engine rooms, tanks and airplanes usually range from 100 to 130 db; hence, in these environments ear wardens should improve speech communication" (**79**). To obtain maximum acoustic insulation, the ear wardens must be properly fitted and those people who wear them must be properly instructed.

Other attempts at providing ear protection have used wax, artificial silk, cotton impregnated with various semi-soft compounds, and even a hood that apparently covered the forehead and cheek bones as well as the ears. Swedish sources (**39, 61**) reported favorably on an isolite ear plug that was tested principally for protection against gun blast. No audiometric data are available on the effectiveness of this plug.

Selection

In most of the laboratory studies as well as many of the clinical observations reported above, a frequent reference has been made to marked individual differences in susceptibility to auditory fatigue and disturbances of functions traceable to high noise levels. It has been suggested that the detrimental effects of noise would be less serious if personnel could be selected who were least susceptible to the adverse effects of noise. In spite of this suggestion, which has appeared in numerous places, there has been little or no attempt to develop screening tests for this purpose. Part of the reason for this deficiency lies in the difficulty of developing reliable measures of annoyance. Investigators have reported that subjective reports of annoyance vary from time to time and from individual to individual in what appeared to be unpredictable ways unrelated to the character of the exposure noise.

It has been suggested further (a) that those people with initial hearing defects might be less susceptible to noise annoyance or (b) that they are more susceptible to noise annoyance. A German study (**68**) reports the effects of exposure to sounds of 100 db for 180 seconds, with audiometric measures taken before and 15 seconds after exposure. The investigator found that ears with chronic inflammation are more sensitive to noise than ears in healthy condition. Thus, persons showing especially a thickening of the mucous membranes on account of chronic middle ear inflammation must be considered especially liable to the detrimental effects of noise.

W. H. Wilson (**98**) suggested a screening technique which involved exposure to a tone of 2048 cycles at 80 db for eight minutes. One minute following exposure a complete audiogram was taken and compared with the pre-exposure audiogram. This was repeated for each ear. Subjects were examined prior to beginning their basic firing on both pistol and rifle ranges. In no instance were the subjects known to have gone out on the range sooner than five days after the above tests. Between the firing and the next audiometric study there was a minimum lapse of 13 and a maximum lapse of 35 hours. It was predicted that those showing an auditory loss of 10 db or greater in the pre-firing tests as a consequence of exposure to 2048 cycles would show a hearing loss on their return from the firing range at a later date. Thirty out of 108 tested fell in this category. Of these, 25 did show such a loss compared with their pre-firing tests. Of the 78 men who did not show initial auditory fatigue, 14 showed a significant decrease in acuity on return from the firing range. Seven days later 26 of the original 39 men who showed hearing losses at the end of firing were retested, and 16 of these still showed some hearing loss (the rest of the original 39 had been sent to other posts).

While this attempt at predicting susceptibility to noise trauma is not definitive, it is sufficiently encouraging to warrant further investigation with exposure to other types of noise and with some variations in basic procedures.

An attempt to use simulated battle noises for the selection of emotionally stable infantry men resulted in the conclusion that such exposures were not an important supplement to other screening devices (**35**). However, since this study was concerned with selecting emotionally stable men and not those least susceptible to noise effects, it cannot be looked upon as a negative finding from the latter standpoint.

One attempt to circumvent the deleterious effects of noise on communication involved the training of pilots to understand voice communication in the presence of noise. Intelligibility scores were compared between initial and final tests with control and experimental groups (experimental groups being exposed to noise inserted in the communication system). In general, practice with noise was advantageous (**14**).

Summary of Research

Many of the studies on the effects of noise, especially those which appeared before 1930, are difficult to evaluate because of unstandardized descriptions of noise levels and frequency characteristics. This is not a reflection on the investigators, but was rather due to the lack of agreement on terminology and inadequate measuring instruments. Unquestionably, over-all noise levels of 110 db and higher are potentially, if not actually, harmful to the auditory functions, especially if the noise is long continued and high frequencies predominate. Although a considerable adaptation takes place with prolonged exposure as far as the non-auditory ill effects are concerned, the evidence suggests but does not fully substantiate the conclusion that this adaptation is not complete. Nearly all studies on the effects of noise have noted individual differences between subjects, making broad generalizations hazardous. In spite of this fact, almost no efforts have been directed toward the objective of screening "noise-prone" individuals, i.e., those most susceptible to noise annoyance or noise trauma.

Areas for Further Investigation

One general area demanding research attention is obviously the selection and screening of those persons least susceptible to noise annoyance and noise trauma, and most readily adaptable to noise conditions. The report of Wilson (**98**) mentioned above is a beginning in this area.

The fact that there are numerous widespread subjective reports of increased general fatigue in the presence of high noise intensities suggests the need for a thorough systematic investigation in this area. What is the basis, physiological or otherwise, of such reports? Measures of annoyance and fatigue have not been entirely satisfactory and need to be developed, if progress in this area is to be made (**45**).

The studies of noise as it affects physical functions have for the most part been deficient both with respect to the levels and kind of noises employed and with regard to the duration of exposure periods. If such studies are to be immediately applicable to problems of submarine warfare, noise levels should vary between 100 and 130 db and the subjects should be repeatedly exposed for continuous periods of ten or more hours' duration. It is possible that the long-run effects on such functions as well as on psychophysical tasks are obscured by long periods of quiet interposed between experimental sessions. Moreover, the factors of motivation and suggestion, both of which are difficult to control, merit more attention than they have generally received in the experimental studies herein reviewed.

The nature and degree of auditory trauma have been relatively well established both experimentally and by clinical observations, so that it is reasonably certain that long exposure to noise levels of 120 db (conditions approximating those in some submarines, airplanes and small surface vessels) will certainly produce a temporary, and probably a permanent, decrease in acuity, especially in the range from 2,000 to 4,000 c.p.s. Although it is beyond the immediate scope of this report, it is also apparent that studies of vibration in the subauditory range and in the ultrasonic range are likewise important. In this connection, D. E. Goldman at the Naval Medical Research Institute, Bethesda, Maryland, is conducting pertinent research.

Finally, a field study of duty assignments aboard submarines may be necessary to discover the extent to which personnel move from high noise levels to assignments demanding high auditory acuity. It may be that a problem of noise protection exists which is comparable to that of dark adaptation of bombardiers or night lookouts. That is to say, men assigned to duties demanding a high degree of auditory acuity should obviously be protected from high noise levels for extended periods prior to taking over their stations. The extent to which such men are not now protected, and the extent to which lack of protection adversely affects performance of duty, need study.

References

1. Atkinson, M. Tinnitus aurium: observations on the effect of curare on loudness level. *Ann. Otol., etc.*, St. Louis, 1946, **55**, 398–405.
2. Baker, K. H. Pre-experimental set in distraction experiments. *J. appl. Psychol.*, 1937, **16**, 471–488.
3. Berrien, F. K. The effects of noise. *Psychol. Bull.*, 1946, **43**, 141–161.
4. Berrien, F. K., & Young, C. W. Effects of acoustical treatment in industrial areas. *J. acoust. Soc. Amer.*, 1946, **18**, 453–457

5. Bogoslovsky, A. I., & Kravkov, S. V. Vliiànie shuma aviatšionnogo motora na zrenie. (The influence of noise of an aircraft engine upon vision.) *Probl. physiol. Opt.*, Acad. Sci. U.S.S.R., 1941, **1**, 69–75.
6. Bunch, C. C. The diagnosis of occupational deafness; an historical and audiometric study. *Laryngoscope*, St. Louis, 1937, **47**, 9–34.
7. Bunker, M. L. Sound levels and noise analysis in engine room of Fairmiles. *Proceedings of 3rd meeting of Associate Committee on Naval Medical Research*, National Research Council of Canada, May 15, 1944. (CONFIDENTIAL)
8. Campbell, P. A. The effect of flight upon hearing. *J. Aviat. Med.*, 1942, **13**, 56–61.
9. Campbell, P. A. The progress of aviation otology in World War II. *Laryngoscope*, St. Louis, 1944, **54**, 387–397.
10. Campbell, P. A. The present status of the problem of aviation deafness. *Trans. Amer. Acad. Ophthal. Otolaryng.*, 1944–1945, **49**, 72–74.
11. Campbell, P. A. The problem of aviation deafness. *Arch. Otolaryng.*, Chicago, 1945, **41**, 319–321.
12. Campbell, P. A., & Hargreaves, J. Aviation deafness—acute and chronic. *Arch. Otolaryng.*, Chicago, 1940, **32**, 417–420.
13. Coermann, R. Lärmeinwirkung im Flugzeng. (Influence of noise in aircraft.) Dtsch. Luftf-Forsch. Forsch Ber No. 1102, 1939; Publ. Bd. No. 37299. Washington, D. C.: U. S. Dep. Commerce, 1946.
14. Curtis, James F. Training studies in voice communication. II. The use of noise in a training program. *OSRD* Rep. No. 4261, Oct. 1944.
15. Davis, H., Derbyshire, A. J., et al. Experimental stimulation deafness. *Science*, 1935, **81**, 101.
16. Davis, H., et al. Final report on physiological effects of exposure to certain sounds. *OSRD* Rep. No. 889, July 31, 1942. (RESTRICTED)
17. Davis, H., Morgan, C. T., Hawkins, J. E., Jr., Galambos, R., & Smith, F. W. Temporary deafness following exposure to loud tones and noise. *Laryngoscope*, St. Louis, 1946, **56**, 1–8.
18. Davis, R. C. Electrical skin resistance before, during and after a period of noise stimulation. *J. exp. Psychol.*, 1932, **15**, 108–117.
19. Davis, R. C. The muscular tension reflex and two of its modifying conditions. *Indiana Univ. Publ.*, Science Series, 1935, No. 3.
20. Duff, Ivan F., Com. (MC)USNR. Medical study of the experiences of submariners as recorded in 1471 Submarine Patrol Reports in World War II. (CONFIDENTIAL)
21. Dyson, G. W. Let's hear about it. *Sci. Mon.*, N. Y., 1946, **62**, 267.
22. Finch, F., & Culler, E. Effects of protracted exposure to a loud tone. *Science*, 1934, **80**, 41–42.
23. Flaiz, J. W., & Wolf, E. G. The effect of noise on aviation personnel. *Contact*, Pensacola, 1943, **3**, 160–167.
24. Fleming, N. Noise and its prevention. *J. Text. Inst.*, Manchr., 1939, **30**, 261–271.
25. Ford, A. Attention-automatization: an investigation of the transitional nature of mind. *Amer. J. Psychol.*, 1929, **41**, 1–32.
26. Freeman, G. L. Changes in tension pattern and total energy expenditure during adaptation to distracting stimuli. *Amer. J. Psychol.*, 1939, **52**, 354–360.
27. Gardner, W. H. Injuries to hearing in industry. *Industr. Med.*, 1944, **13**, 676–679.
28. Goldner, A. Occupational deafness. *Arch. Otolaryng.*, Chicago, 1945, **42**, 407–411.
29. Grabener, H. Auditory deterioration in airline pilots. *J. Aviat. Med.*, 1947, **18**, 39–47.
30. Grossman, F. M. Sound transmission through the ear and its relation to sound injury. *Ann. Otol., etc.*, St. Louis, 1943, **52**, 666–673.
31. Hagerman, F. Hearing injuries after acoustic shot traumata. *Acta oto-laryng.*, Stockh. 1942, **30**, 75–86.
32. Harmon, F. L. The effects of noise upon certain psychological and physiological processes. *Arch. Psychol.*, 1933, No. 147.
33. Harris, J. Donald. Pitch discrimination in noise. *Navy Dept.*, Med. Research Lab., U. S. Submarine Base, New London, Conn., Interim report No. 1, July 6, 1945. (SECRET)
34. Harris, J. D., & Stover, A. D. Noise levels aboard a fleet submarine. Progress Report No. 2, Bureau of Med. and Surgery, Research Project NM-003-020, U. S. Submarine Base, New London, Conn. April, 1948. (RESTRICTED)
35. Hartley, E. L., & Jones, D. B. The selection of emotionally unstable personnel by battle noise equipment. *OSRD*, 1945, Washington, D. C., U. S. Dept. Commerce, 1946, Pub. Bd. No. 12052. (RESTRICTED)
36. Hendricks, J. E. War deafness in gunnery instructors. *Ann. Otol., etc.*, St. Louis, 1946, **55**, 68–80.

37. Heyden, R. Aviators' deafness. *Acta oto-larying.*, Stockh. 1944, **32**, 164–175.
38. Hipskind, M. M. War injuries to the auditory apparatus; clinical and audiometric observations. *Laryngoscope*, St. Louis, 1945, **55**, 678–705.
39. Holmgren, L. Some general and special remarks on the investigations regarding aural protections in the army during the autumn of 1941. *Acta oto-laryng.*, Stockh. 1944, **32**, 502–509.
40. Hoople, Gordon D., Wolfe, William C., & Bregande, Samuel C. Unrecognized battle noise trauma. *Laryngoscope*, St. Louis, 1947, **57**, 125–136.
41. Howard, J. C. Hearing—A post-war problem. *Ann. Otol., etc.*, St. Louis, 1943, **52**, 843–849.
42. Johnson, N. C. Ear disorders in flying. *Contact*, Pensacola, 1947, **6**.
43. Katz, B. An electric "reaction meter" suitable for prolonged performance tests. *Australian Technical Papers*, No. 719, Report No. 3, Acoustic Testing Lab., Sydney.
44. Kos, C. M. Effect of barometric pressure changes on hearing. *Trans. Amer. Acad. Ophthal. Otolaryng.*, 1944–1945, **49**, 75–81.
45. Kryter, K. D., & Reese, T. W. The relative annoyance produced by various bands of noise. Informal Communication No. 65, *NDRC*, March 1944. (RESTRICTED)
46. Laird, D. A. Experiments in the physiological cost of noise. *J. nat. Inst. ind. Psychol.*, 1929, **4**, 251–258.
47. Laird, D. A. Influence of noise on production and fatigue, as related to pitch, sensation level and steadiness of the noise. *J. appl. Psychol.*, 1933, **17**, 320–330.
48. Laird, D. A., & Coye, K. Psychological measurements of annoyance as related to pitch and loudness. *J. acoust. Soc. Amer.*, 1929, **1**, 158–162.
49. Larsen, B. Investigations in the fatigue of hearing. *Acta oto-laryng.*, Stockh., 1942, **30**, 525.
50. Lindahl, R. Noise in industry. *Ind. Med.*, 1938, **7**, 664–669.
51. Loye, D. P., Wagner, R. A., & Pratt, R. W. Overside noise measurements of submarines. *OSRD* Rep. No. 4223, September 12, 1944. (CONFIDENTIAL)
52. Lurie, M. H. War and deafness. *Conn. med. J.*, 1943, **7**.
53. McCoy, D. A. The industrial noise hazard. *Arch. Otolaryng.*, Chicago, 1944, **39**, 327–330.
54. Machle, W. Effects of gun blasts on hearing. *Arch. Otolaryng.*, Chicago, 1945, **42**, 164–168.
55. McKelvie, W. B. Weaver's deafness. *J. Laryng., Rhin. Otol.*, 1933, **48**, 607.
56. Malone, P. W. Aviation deafness. *Arch Otolaryng.*, Chicago, 1944, **40**, 468–474.
57. Maninger, R. C. Experimental investigation of factors involved in sonic listening. *OSRD*, 1945, Washington, D. C., U. S. Dep. Commerce, 1946, Pub. Bd. No. 33397. (CONFIDENTIAL)
58. Maninger, R. C., & Knudson, W. T. Tests on 1MC and 7MC intercommunication systems of the USS PERCH. Columbia University Division of War Research, Feb. 17, 1944. (CONFIDENTIAL)
59. Merkel, B. M. Aviation deafness. *J. Ia. St. med. Soc.*, 1946, **36**, 288–291.
60. Morgan, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol.*, 1916, No. 35.
61. Nasiell, V. Principles concerning the prophylaxis of aural injuries in the services. *Acta oto-laryng.*, Stockh., 1944, **32**, 509–526.
62. Olmsted, J. M. D. The special senses. *Ann. Rev. Physiol.*, 1945, **7**, 509–526.
63. Parkinson, J. S., & Jack, W. A. Reexamination of the noise reduction coefficient. *J. acoust. Soc. Amer.*, 1941, **13**, 163–169.
64. Patt, H. M. Hearing loss from gunfire. *Hosp. Cps. Quart.*, Washington, D. C., 1945, **18**, No. 11, 41–42.
65. Patt, H. M. Auditory acuity of artillery men and of rifle coaches. *Arch. Otolaryng.*, Chicago, 1946, **44**, 298–302.
66. Pedley, F. G. Incidence of occupational deafness and methods for its prevention. *Safety Eng.*, 1930, **60**, 237.
67. Perlman, H. B. Acoustic trauma in man. *Arch. Otolaryng.*, Chicago, 1941, **34**, 429–435.
68. Peyser, A. Zur Methodik einer otologischen Prophylaxis der industriellen Lärmschwerhörigkeit. 2. Bericht. (On the methodology of otologic prophylaxis of hearing deficiency in industry due to noise. Second report.) *Acta oto-laryng.*, Stockh., 1943, 31, 351–364.
69. Pollock, K. G., & Bartlett, F. C. Psychological experiments on the effect of noise. *Indus. Health Res. Bd.*, 1932, Rep. 65.
70. Rasmussen, H. Studies on the effect upon the hearing through air conduction brought about by variations of the pressure in the auditory meatus. *Acta oto-laryng.*, Stockh., 1946, **34**, 415–424.
71. Rettaliata, J. T., & Green, W. P. Reduction of noise in light-weight high-speed

gasoline engines. Med. Eng. Dep., Ill. Inst. of Tech., Chicago, Ill. and U. S. Army Signal Cps. Eng. Lab., Bradley Beach, N. J., November 21, 1947.

72. Roger, L. R. Noise-deafness: a review of recent experimental work and a clinical investigation into the effect of loud noises upon the labyrinth in boilermakers. *J. Laryng., Rhin., Otol.*, 1915, **30**, 91–105.

73. Rosenblith, W. A. Industrial noises and industrial deafness. *J. acoust. Soc. Amer.*, 1942, **13**, 220–225.

74. Scheiwechmann, B. E. Der Einfluss der Schallwellenfrequenz auf das Gehör der Arbeiter in mit Lärm behafteten Betrieben. (The influence of the frequency of sound waves on the hearing of workers in noisy jobs.) *Bull. Biol. Med. exp.*, U.S.S.R., 1939, **8**.

75. Seashore, H., & Stuntz, S. Training radio operators to copy code through interference. *OSRD* Rep. No. 3747, June 7, 1944.

76. Senturia, B. H. Auditory acuity of aviation cadets. *Ann. Otol., etc.*, St. Louis, 1944, **53**, 705–716.

77. Senturia, B. H. The effect of aircraft noise upon hearing. *Trans. Amer. Acad. Ophthal. Otolaryng.*, 1944–1945, **49**, 82–89.

78. Senturia, B. H. Effect of aircraft noise on hearing. *Arch. Otolaryng.*, Chicago, 1945, **41**, 327–332.

79. Shaw, W. A. Development of ear wardens type V-51R. *OSRD*, 1945, Washington, D. C.: U. S. Dep. Commerce, 1946, Pub. Bd. No. 22849.

80. Shilling, C. W. Aero-otitis media and auditory acuity loss in submarine escape training. *Trans. Amer. Acad. Ophthal. Otolaryng.*, 1944–1945, **49**, 97–102.

81. Shilling, C. W. Aero-otitis media and loss of auditory acuity in submarine escape training. *Arch. Otolaryng.*, Chicago, 1945, **42**, 169–173.

82. Shilling, C. W., & Everley, I. A. Auditory acuity among submarine personnel. *Nav. med. Bull.*, Washington, D. C., 1942, **40**, 27–42.

83. Silcox, L. E., & Schenck, H. P. Blast injury of the ears. *Arch. Otolaryng.*, Chicago, 1944, **39**, 413–420.

84. Smith, E. L., & Laird, D. A. The loudness of auditory stimuli which affect stomach contraction in healthy human beings. *J. acoust. Soc. Amer.*, 1930, **2**, 94–98.

85. Stevens, S. S. The effects of noise and vibrations on psychomotor efficiency. *OSRD*, 1941, Washington, D. C., U. S. Dep. Commerce, 1946, Pub. Bd. No. 8333.

86. Stevens, S. S., et al. Progress report on Project II, Dec. 2, 1941: I. The effects of noise on psychomotor efficiency. II. Noise reduction in aircraft as related to communication, annoyance, and aural injury. *OSRD*, 1941, Washington, D. C., U. S. Dep. Commerce, 1946, Pub. Bd. No. 8334.

87. Stewart, J. V., & Barrow, D. W. Concussion deafness. *Arch. Otolaryng.*, Chicago, 1946, **44**, 274–279.

88. Suggit, S. Deafness resulting from gunfire and explosions. *J. Laryng. Otol.*, 1943, **58**, 313–326.

89. Swann, C. C. The effect of noise on hearing. *Internat. J. Med. Surg.*, 1933, **46**, 314–315.

90. Ullman, E. V. Traumatic deafness in combat flyers. *Arch. Otolaryng.*, Chicago, 1944, **40**, 374–381.

91. Vadala, A. J. Effects of gun explosion on the ear and hearing mechanisms. *Mil. Surgeon*, 1930, **66**, 810–822.

92. Vernon, H. M., & Warner, G. G. Objective and subjective tests for noise. *Personnel J.*, 1932–1933, **11**, 141–149.

93. Weiss, J. A. Deafness due to noises of warfare. *Nav. med. Bull.*, Washington, D. C., 1946, **46**, 381–386.

94. Weston, H. C., & Adams, S. The effects of noise on the performance of weavers. *Indust. Hlth. Res. Bd.*, 1932, Rep. 65.

95. Weston, H. C., & Adams, S. The performance of weavers under varying conditions of noise. *Indust. Hlth. Res. Bd.*, 1935, Rep. 70.

96. Wever, E. G. Audition. *Ann. Rev. Physiol.*, 1946, **8**.

97. Wilson, A. Better concentration reduces employee turnover 47 percent, adjustment of noise factor increased general efficiency 8.8 percent. *Bankers Mo.*, 1942, **59**, 254–255.

98. Wilson, W. H. Prevention of traumatic deafness. *Arch. Otolaryng.*, Chicago, 1944, **40**, 52–59.

99. Winfield, R. H. Factors influencing onset and production of fatigue in Catalina Flying Boat Crews. *Flying Personnel*, Canada, 355. (CONFIDENTIAL)

100. Investigation of ear plugs for protection against gun blast. NDRC-ear wardens, M.S.A. ear defenders, SEPCO ear protectors, cotton, rubber plugs with baffles and nods. *Armored Med. Res. Lab.*, Fort Knox, Ky., August 5, 1944.

Chapter 17

LIGHTING AND COLOR

MILES A. TINKER

University of Minnesota

Introduction

Lighting

The problem of artificial illumination is of primary importance in all inside living and working environments including the submarine. To maintain healthful and efficient functioning of the eyes, it is necessary to provide adequate lighting. Unquestionably, proper illumination contributes much to comfort and efficiency in activities of daily life. Performing visual work under faulty illumination frequently results not only in visual discomfort, but in eyestrain which tends to be accompanied by reflex disturbance of other organs of the body.

In the submarine only artificial illumination is used, and members of the crew have normal vision. Accordingly it is necessary to consider those factors which are basic to *artificial* illumination adequate for maintaining comfortable, healthful and effective functioning of *normal* eyes. These factors are: (1) quality or color of light; (2) intensity of light in relation to (a) visual acuity, (b) size of object to be discriminated, (c) speed of vision, (d) brightness contrast, (e) efficiency of performance; and (3) distribution of illumination and brightnesses within the working environment. The latter includes various forms of glare and brightness contrast. This distribution of illumination should be so arranged as to give compartments more pleasing appearance and proportions. All three factors should be combined to induce in the inhabitant a sense of tranquillity, comfort and "hominess."

The literature on illumination is enormous. This report will, therefore, be based upon a sampling of this literature.

Color

Certain colors and tints of colors are preferred over other colors and tints. Some colors carry the meaning of warmth, others appear cool. Some are exciting, others subduing. To achieve the most pleasing living and working environment, these psychological factors must be considered.

The role that color in decoration (paint) plays in promoting adequate illumination is important. Much of the usable illumination in an enclosed space should come from reflection of light from the ceiling, walls and furnishings. Consequently, attention must be given to the reflection factor of the paint. Also one must consider coordination of illumination and decoration.

Spectral Quality of Color of Light

In specifying artificial illumination one must decide what light should be used. Several illuminants, varying in spectral character, are available: artificial daylight, tungsten-filament incandescent light, mercury arc light, sodium vapor light, and fluorescent light.

It is well known that the lens in the human eye has a slightly different focal length for each color of the spectrum. This results in chromatic aberration when the light reflected from an object to the eye is complex, i.e., involves several wave lengths. For very near vision, it is possible to focus for violet rays, but somewhat difficult to focus for red rays. In far vision the reverse is true. For intermediate distances it is easiest to focus for yellow rays. Thus, when the object is illuminated by mixed wave-lengths as in ordinary lighting, we have chromatic aberration in which the violet rays come to a focus farther to the front of the eye than the red rays, and the yellow rays occupy an intermediate position. These intermediate rays are automatically focussed upon the retina. This results in aberration circles for the red and violet which produce a slight blurring of the optical image, and consequently a reduction in visual acuity. This blurring is not noticeable in ordinary circumstances because the brightness of the yellow rays is much greater than that of the violet or the red. Furthermore, the eye is much more sensitive to yellow than to red or violet. Consequently the aberration circles are seldom noticed. We might expect, however, that vision would be clearer if all rays but those from one region in the spectrum were eliminated. That is, the more nearly monochromatic a light, the sharper the image on the retina.

Comparison of Various Illuminants

Visual acuity varies measurably from one monochromatic light to another. Ferree and Rand's careful determinations (**10**) found the order yielding greatest to least acuity to be: yellow, yellow-green, orange, green, red, blue-green and blue. But sunlight was superior to any of the colors.

Sodium versus Tungsten Light

Luckiesh and Moss (**16**) measured visual acuity under sodium vapor and under tungsten filament light. For all practical purposes, the sodium light was monochromatic. The brightness range was from 0.018 to 20 millilamberts. Acuity of nine subjects was measured by means of the Ives-Cobb acuity object. At all levels of brightness, the sodium light produced greater acuity. Speed of visual reaction was found, in the same study, to be greater under sodium than under tungsten light when the test object is small, the brightness one millilambert, and the contrast between object and ground large. But when the brightness was only 0.1 millilambert, the test object large, and the brightness contrast very small, there was no difference in speed. Incidentally, it has been demonstrated by Luckiesh and Holladay (**14,** p. 246) that sodium yellow light does not penetrate fog appreciably better than white light. It is apparent from these studies that where very fine (threshold) discrimination is concerned, sodium vapor light is slightly better than ordinary tungsten. However, as demonstrated by Luckiesh (**14,** p. 246), this advantage for sodium light or any monochromatic light disappears when objects to be discriminated are well above threshold size. For example, visibility of book print (8 to 10 point) is about the same under sodium and under tungsten light. Sodium light is undesirable for general use because all color but yellow disappears. It has no advantage in enhancing contrasts (**14,** p. 246).

Mercury Arc Light versus Tungsten Filament Light

It was commonly held that in view of the sharp retinal image obtained with monochromatic light (as referred to above), visual acuity would be improved the nearer one gets to monochromatic illumination. (It might be pointed out that some people consider that there is no rational argument for this hypothesis.) Therefore, mercury arc light, which has relatively homogeneous wave lengths, should produce greater visual acuity than tungsten filament light. Luckiesh and Moss (**18,** p. 406) question the validity of early experiments on this subject. They (**17**) measured visual acuity, under 5, 25, and 125 foot-candles of illumination from the low-pressure mercury arc and a gas filled tungsten-filament lamp. Measurements were obtained from 10 subjects on the Ives-Cobb acuity object. At each foot-candle level there were only insignificant differences between the lights in visual acuity. They conclude that the mercury light produces no greater visual acuity because of less chromatic aberration than the tungsten light which has a continuous spectrum.

In another experiment, Luckiesh and Moss (**18,** p. 418) report that the visibility of book print is the same under mercury arc and under tungsten light. The visibility measurements, made on eight subjects for 8-point book type by means of the Luckiesh-Moss Visibility Meter, averaged 3.8 for the mercury arc and 3.7 for the tungsten light. Other data (**18**) confirm the above findings. In general, these results indicate that the mercury arc light is no better and no worse than tungsten filament for visual work.

Fluorescent versus Tungsten Filament Light

Periodically, since the advent of fluorescent lighting, questions have been raised concerning possible harmful effects to the organism. Luckiesh and Taylor (**21**) have evaluated the criticisms of fluorescent lighting. They present evidence which leads them to conclude that the ultraviolet, infrared, and visible radiant energy coming from the lamps is not injurious to the human organism. This has not quieted either the criticisms of fluorescent light or the complaints of those working under it (**12**). Apparently fluorescent is as effective as tungsten filament light for visual efficiency. Luckiesh and Moss (**20**) show that the two are equally effective in promoting ease of seeing. The change in rate of involuntary blinking after reading for 30 minutes was the same for both kinds of illumination.

Holway and Jameson (**12**) carried out extensive investigations in which one of the comparisons was between fluorescent and incandescent illumination. The quality of light derived from the fluorescent lamps, no matter what combination of colors was used, was considered both unpleasant and distracting to the workers. The quality was described as thin, harsh, and cold. This undesirable quality was not present in either daylight or incandescent illumination. This psychologically undesirable quality of light from fluorescent lamps was considered a factor detrimental to ease and comfort in reading. The authors recommend that incandescent rather than fluorescent light be used in reading rooms and offices.

These authors (**12**) also state that experimentation has shown that, after working under direct fluorescent illumination, the recovery of sensitivity in the night vision apparatus is retarded seriously in comparison to being under incandescent light. It is stated that this effect is caused by ultraviolet radiation from fluorescent light sources. After 30 minutes of dark adaptation, following an hour under fluorescent light, night vision was only half as good as it was after being under incandescent light. This retardation in recovery of night vision can be eliminated when fluorescent light is shielded or used in indirect fixtures.

Summary

For threshold seeing, spectral yellow, or yellow light from sodium vapor lamps, is slightly more effective than other artificial illuminants. But these differences are small and relatively inimportant. Apparently, no illuminant is better than diffused daylight. In ordinary situations, where seeing is supraliminal, as in reading print, no artificial illuminant that is suitable in spectral character for general use in lighting is better than any other such illuminant for critical seeing. This is true for daylight, and for light from mercury arcs, tungsten filaments, or fluorescent lamps. There is, however, a psychological objection to fluorescent light because of its "harsh" and "cold" appearance.

Intensity

In any environment where visual work is to be done, one must decide how much light (i.e., what intensity of illumination) is needed for easy and effective visual discrimination. Various approaches have been employed to arrive at proper light intensities for effective seeing. We will discuss here the relation of illumination intensity to (1) visual acuity; (2) size of object to be discriminated; (3) speed of vision; (4) brightness contrast; and (5) efficiency of performance. An appraisal will be given after presenting the data in each area.

Visual Acuity

In considering the relation between illumination and visual acuity, one is dealing with the discrimination of fine details, or threshold vision. The customary method is to determine the least visual angle at which the direction of the break in the Landolt ring (international test object) can be perceived for a given light intensity. Usually five out of eight correct responses are required. Intensity is usually measured in foot-candles. One foot-candle (f.c.) is the light intensity of a standard candle at the distance of one foot. The data presented by Feree and Rand (**8**) show that when the intensity is increased from .001 f.c., the rise in visual acuity is rapid up to an illumination of about 5 f.c. As the intensity is increased further the rise in acuity becomes progressively slower and practically reaches a maximum at about 20 f.c. Improvement in the later stages is scarcely noticeable. The varying effects of intensity increase at different levels of illumination is revealed by a 4.89-fold increase in acuity from .001 to 0.1 f.c., and 67.7 percent increase from 0.1 to 1.0 f.c., a 43.6 percent increase from 1.0 to 5.0 f.c., and a mere 8.2 percent increase from 5.0 to 20.0 f.c. In a later study, Feree, Rand and Lewis (**11**) employed illuminations ranging from 0.5 to 100 f.c. For their subjects with normal vision, the increase of visual acuity was rapid up to 10 f.c., small from 10 to 25 and very slight from 25 to 100 f.c. Normal eyes receive slight benefit from increases beyond 25 f.c.

One of the most painstaking studies of the relation between visual acuity and illumination is reported by Lythgoe (**23**). The illumination used ranged from 0.0029 to 1275 equivalent foot-candles (e-f.c., i.e., foot lamberts). Repeated measurements were made on four subjects. The reflection factor of the test object (Landolt ring) was 1.35 percent and of the background, 71 percent. As in the earlier studies, the increase in visual acuity was rapid up to about 5 e-f.c., slow to 38, and slight to 1275. Acuity was raised only about one-tenth of a point (2.10 to 2.24) when the illumination was raised from 38.8 to 208 e-f.c., and another tenth (2.24 to 2.35) when raised to 1275 e-f.c. Employing the Ives-Cobb acuity object, Luckiesh and Moss (**17**), using 10 subjects, measured visual acuity under 5, 25, and 125 f.c. of mercury-arc light and of tungsten-filament light. Although there was a marked increase in acuity as the light was changed from 5 to 25 f.c., the slight improvement from 25 to 125 f.c. was not significant.

The representative data on visual acuity in relation to illumination cited here reveal the following trends: (1) when the illumination intensity is increased from a fraction of a f.c. to about 5 f.c. there is a rapid increase in visual acuity. (2) From 5 to 20 f.c. there is a gradual increase in acuity; (3) after a level of 25 f.c. is reached, the rate of increase in visual acuity is small. The curve of improvement is practically flat from about 40 f.c. on. Although slight increases in visual acuity can be obtained by higher illumination, they are of no practical importance. Thus, 170 e-f.c. must be added to 38 to get an increase in visual acuity of one-tenth of a point. An additional 1000 e-f.c. are needed to increase acuity another tenth of a point. In general, with normal eyes, little is gained in acuity by increasing the illumination beyond 25 f.c., and there is no practical gain at all when the intensity is higher than approximately 50 f.c.

Many writers, including Lythgoe (**23**), point out that the relationship between visual acuity and illumination intensity is logarithmic. Expressing this relationship on a logarithmic scale, however, magnifies minute acuity differences at the higher levels of illumination so that they appear *large* and *significant*. This is especially true when one is thinking of practical applications of illumination in daily life. There are limitations to prescribing light on the basis of the relation of illumination to visual acuity. These limitations will be considered later (under *Brightness Contrast*).

Size of Object

The smallest size of detail that can be barely seen under a given set of conditions is the threshold size. The most complete study on the relation between size and illumination was done by Weston (**33**). Eighteen subjects cancelled all the Landolt rings with a given gap orientation on specially prepared test sheets. The rings were black, printed on paper with a reflection factor of 84 percent. The gap size at 13 inches from the eye subtended (on different sheets) angles of 1, 2, 3, 4, 6, and 10 minutes. The detail (space between dot and rest of letter *i*) of 10 point type (ordinary book type) subtends an angle at the eye of approximately 3 minutes at a distance of 13 inches. Illumination intensities used were 0.16, 0.8, 4, 20, 100 and 500 f.c. Both accuracy and speed of discrimination entered into the scores. The results are given in terms of speed (reciprocal of time), which is unfortunate, since this magnifies small differences unduly. The results show (a) that for all sizes of objects there was a rapid increase in performance as the light was increased from 0.16 to 4.0 f.c., and (b) that for objects of 4 minutes visual angle and larger there was no appreciable increase for illuminations above 20 f.c. For small objects (1, 2, and 3 minutes), the improvement of visual performance is small and probably not significant beyond 50 f.c., contrary to the author's view. It is pointed out that no practical amount of illumination will make small objects (1 minute) as readily discriminated as large objects (6 minutes).

Luckiesh (**14,** p. 98) also presents data on the relation between size of object and illumination in seeing. As the illumination is increased from 1 to 10 to 100 f.c., smaller and smaller sized objects can be discriminated for all degrees of contrast between object and ground. The limits of practical improvement undoubtedly lie between 10 and 100 f.c. Interpolation reveals that the curve of improvement flattens out rapidly for levels above 30 to 40 f.c.

These data reveal (a) that for large objects (4 minutes and above), there is no practical improvement in visual discrimination with intensities above 20 f.c. For smaller objects there is improved visual discrimination at higher intensities. The indications are, however, that these changes are significant only up to 40 or 50 f.c. As a matter of fact, these data merely represent another way of measuring visual acuity. The amount of illumination needed to perceive an object also depends upon the contrast between object and background. This will be treated below (see *Brightness Contrast*).

Speed of Vision

It takes time to see. Quickness of seeing is influenced by level of illumination as well as by size of object to be discriminated and brightness of the background. Where split-second seeing is desirable, the time factor is important. Luckiesh (**14,** p. 131) reports data which show that the time required to recognize a black test object decreases markedly as the brightness-level of the background increases from 1 to 2 to 5 to 10 footlamberts. (One footlambert is the brightness of a perfectly reflecting surface illuminated by one f.c.) As the brightness of background is further increased to 20 footlamberts, the additional decrease in time is about .005 second, and there is a like decrease in going from 20 to 100 footlamberts. Thus, the decrease in time of seeing is relatively insignificant for the more intense illumination changes. For a small test object, Cobb (**5**) found no significant decreases in speed of vision for brightnesses above about 40 footlamberts when a pattern of parallel dark bars was presented before and after the test object, i.e., a situation comparable to seeing in everyday life where the eye receives stimulation before and after discrimination of an object. The range of illumination used was 1 to 100 footlamberts. With a relatively large test object and similar conditions, there was no significant decrease in time for brightnesses above about 18 footlamberts.

The most extensive study of speed of vision was done by Ferree and Rand (**9**). They investigated the influence of illumination on speed of vision with test objects of 1, 2, 3, 4.2, and 5.2 minutes visual angle and with backgrounds of 78, 29, 21 and 16 percent reflection. The test object was black, reflecting only 3 to 4 percent. Illumination intensities ranged from 1.25 to 100 f.c. The standard Landolt ring was used as test object. In all cases the time taken to perceive the test object decreased rapidly up to 30 or 40 f.c. (up to the knee of the curve of improvement) for the small test objects (1 and 2 minutes), and up to 15 to 25 f.c. for the larger test objects (3, 4.2 and 5.2 minutes). In all cases there was a slower decrease for about 5 to 10 f.c. beyond the knee of the curve and then a rather rapid flattening out of the curve. With the lower degrees of background brightness, the improvement in speed extended to higher levels of illumination. These data signify that time of seeing decreases by significant amounts up to about 20 or 30 f.c. for the larger test objects and up to 40 or 50 f.c. for the smaller test objects. The reciprocal of the time for perception was used for plotting the curves of change of speed with increased illumination intensity. This magnifies slight differences so they look appreciable in the figures. For instance, the decrease in time of perceiving test objects 2 to 5.2 minutes in size as the illumination is increased from 20 to 100 f.c. is only 1.5 ms. (thousandths of a second); for the 1-minute test object, the decrease in time as the light is changed from 50 to 100 f.c. is about 7 ms. Changes of this magnitude are of no practical significance. Arnold and Tinker (**1**) have demonstrated that the shortest average pause of the eyes to discriminate an object is 157 ms. In reading, the average pause ranges from about 200 to 300 ms. They also cite data which indicate that when visual impressions succeed one another, as in everyday life situations, an exposure of 100 ms. is necessary for a fairly well cleared up visual impression. These data indicate that any decrease of a few ms. is of no practical importance, since the impression (stimulus) must act for at least 100 and more likely for 200 or more ms. for an adequate visual impression.

The data on speed of vision may be summarized as follows: (a) for objects that are 3 minutes in size or larger and with good contrast between object and background, speed of vision is near maximum at about 15–20 f.c.; (b) for small objects on a background of low reflectance, significant decreases in time for seeing occur with illumination intensities up to about 50 foot-candles; (c) the findings agree with those on visual acuity.

Brightness Contrast

It would be impossible adequately to specify lighting for a specific situation in terms of the relation of illumination to visual acuity, size of object, or speed of vision. In each of these the brightness contrast between the object to be discriminated and the background is very important. In any seeing situation, the inter-relation of three factors determine ease of seeing: size of details, brightness contrast between details and background, and level of illumination on the task. In general, the greater the brightness contrast, the better the seeing at any level of illumination. Luckiesh (**14**) presents data which demonstrate that vision improves as the brightness contrast changes from 2.5 to 12.5, 25, 50 and 96 percent, whether 1, 10 or 100 foot-candles of light are used.

The most complete investigation of illumination and the effects of brightness contrast was done by Weston (**34**). The task required the discrimination and cancellation of Landolt rings with a given gap orientation on specially prepared test sheets. There were 15 subjects in Part I and 12 in Part II. In Part I, the percent contrasts were 36.5, 68.3, and 91.6; the size of test object 1, 3 and 6 minutes (visual angle); and illumination levels 0.8, 4, 20, 100, and 500 f.c. Both accuracy and speed are included in the performance score. For the 6-minute object, performance varied little over a wide range of illumination values for the two higher brightness contrasts. And for 36.5 percent contrast, the improvement is slight above 20 f.c. For the 3-minute size of object, performance is nearly constant for 20 f.c. and above, with the two higher contrasts; but it increases significantly through 20 but not beyond 100 f.c. for the lower contrast.

For the 1-minute size of detail all performance is on a lower level. Performance rises rapidly through 20 foot-candles and thereafter the improvement becomes slower and slower. Apparently the optimum practical illumination level for 1-minute size is between 20 and 100 foot-candles. It should be noted that, although lower brightness contrast brings poorer performance at each illumination level, increasing the illumination intensity actually becomes less effective in improving speed of performance as the brightness contrast becomes smaller. In Part II of the experiment, sizes of test obfects were 1.5, 3.0 and 4.5 minutes. The illumination levels were 0.5, 2, 8, 32, 128 and 512 f.c. The findings are approximately the same as in Part I. The improvement leveled out at around 32 or 128 f.c. With high contrast (97 percent) the leveling off occurred at 8 and 32 f.c. In 3 out of 13 instances it was between 32 and 128. There were slight improvements which were insignificant from a practical viewpoint up to 512 f.c. Again brightness contrast was potent in determining efficiency of performance.

The data on brightness contrast may be summarized as follows: (a) When brightness contrast between object and background is high, discrimination of 3- to 6-minute sizes is not significantly improved by illumination above about 20 f.c. (b) With small size (1-minute) object, performance improves significantly with increased light intensity up to about 30 f.c. or to a level between 20 and 100 (or 32 and 128) f.c. The curves suggest that the highest practical illumination is around 50 to 60 f.c. (c) The greater the brightness contrast, the better the visual discrimination at any illumination level. (d) Excessive illumination will not compensate for small object size or for poor contrast. (e) Nevertheless, slight improvement is achieved up to 500 f.c. for discriminating smaller object sizes (1 to 3 minutes) and objects with poor brightness contrast.

Efficiency of Performance

An analysis of various criteria employed for specifying light intensities has been made by Tinker (**31**). Efficiency of performance, although not an ideal criterion, is more satisfactory than certain others, which appear to lack validity. The earlier data on the relation of illumination intensity to efficiency of performance are summarized by Tinker (**30**).

Critical Levels of Illumination

The critical level of illumination is the intensity beyond which there is no further increase in efficiency of performance as the foot-candles become greater. In Tinker's summary (**30**) the following critical levels of illumination are listed: for reading legible print (about 10 point on good paper) by adults, it is approximately 3 to 4 f.c.; for reading and study of children, 4 to 6 f.c.; for arithmetical computations, less than 9.6 f.c.; for sorting mail, 8 to 10 f.c.; for the exacting task of setting 6-point type by hand, 20 to 22 f.c.; for the very fine discrimination required to thread a needle, 30 f.c.; and for reading newspaper print, about 7 f.c.

Other Studies on Illumination Level

After an extended period of experimentation, Holway and Jameson (**12**) arrived at 20 f.c. as the amount of light needed on the work surface in offices and reading rooms. An intensive study on effects of illumination level on visual performance has been completed by Simonson and Brozek (**27**). Six subjects were used. The task was to discriminate small letters (3.5 points) as they passed a window. Illumination intensity ranged from 2 to 300 f.c. There was a significant increase in performance up to 50 f.c.

Then there was a further slight increase at 100 f.c., but it was not significant. From 100 to 300 f.c. there was a decrease in performance. The authors interpret this to mean that 100 f.c. is optimum for exacting visual work. Actually, the increases in performance beyond 50 f.c. were neither significant nor of practical importance. It is noteworthy that the authors found a falling-off in discrimination at the higher levels. Luckiesh and Moss (**18**, p. 205) discovered that the nervous muscular tension at subjects' finger tips as they were reading was less under 10 than under 1 f.c., and less under 100 than under 10 f.c. of light. Reanalysis (**31**) of the data indicated that rather rapid changes in tension occurred from 1 to about 10 f.c., smaller changes from 10 to around 20 to 25 f.c., and only slight changes from there on to 100 f.c.

Margin of Safety

It is obvious that visual work should not be done at critical levels of illumination. There should be an adequate margin of safety to provide for individual variation and the like. The addition of about 10 f.c. to the critical level will ordinarily provide an adequate margin of safety.

Adequate Foot-Candles

Taking into account a margin of safety, performance studies indicate the following as adequate illumination: (a) 15 to 20 f.c. for reading 10- to 11-point type on good white paper; (b) 20 to 25 f.c. for reading newsprint; (c) 25 to 35 f.c. for discriminating 6-point type. The more severe visual tasks of daily life apparently require 40 to 50 f.c.

Summary on Intensity

When considering the intensity of illumination in relation to visual acuity, size of object, speed of vision, brightness contrast, and efficiency of performance, the composite picture is as follows: visual efficiency increases rapidly up to about 5 f.c., more slowly to 10, very slowly to 20, and by insignificant amounts thereafter, when the object to be discriminated is of moderate size (3 to 6 minutes). But when the object is small, vision improves by practical amounts up to 40 or 50 f.c. The greater the brightness contrast, the better the visual efficiency. Although not practically significant, visual acuity and speed of vision continue to improve slightly up through and beyond 100 f.c. Certain authors (**18, 34**) claim these slight gains at high intensities are important. An evaluation of these claims is given by Tinker (**31**).

Distribution of Illumination

In practice, the most fundamental aspect of healthful illumination, and the one most frequently inadequate, is the distribution of light and brightness in the field of vision. For convenience the discussion of distribution may be considered under glare and brightness contrast. When light enters the eyes from any visible light source so that vision is impaired or discomfort and annoyance is felt, or both things occur, glare is present. Brightness contrast refers to the variation of brightness of areas within the visual field.

Glare

Everyone is familiar with the disagreeable effects of a bright side light or reflection from a highly polished piece of metal shining into the eyes while reading, doing other visual work, or even when no exacting visual discrimination is involved. Cobb and Moss (**6**) studied the effect of a glare source on visibility. The glare source was a 100 watt bulb located at 40, 20, 10 and 5 degrees from the line of vision and in the vertical meridian. The bulb, located 55 inches from the eyes, produced five f.c. of glare light at the eyes. Visibility was reduced appreciably with the glare source at 40 degrees from the line of vision. This loss became progressively greater as the glare source was brought closer to the line of vision. At the 5-degree position, 84 percent of the illumination on the test object was wasted by glare. The effects of such glare can be largely eliminated by increasing the brightness contrast of test object and its immediate background and by increasing the illumination on the test object. *It is better, however, to eliminate the glare source.* Ferree and Rand (**7**) have demonstrated that after reading for four hours under three systems of lighting, the greatest loss of efficiency occurred with the direct, next greatest with the semi-indirect, and least of all with the indirect system. Where several light sources are in the visual field, the greater the number, the greater the loss of visual efficiency. Luckiesh (**14**, p. 204) shows that nervous muscular tension is much greater when a glare source is in the field of view.

Specular reflection, as reflection from a mirror or shiny metal, is also a glare source. Reflections from glossy or polished surfaces, when light is not well diffused, produce visual discomfort and may lessen visual efficiency. Such surfaces should be eliminated. Luckiesh (**14**) shows that in the case of printed matter, specular brightness may be greatly reduced by increasing the light intensity. A better solution is to use well-diffused illumination.

Brightness Contrast

The distribution of brightness in the visual field is an important determinant of seeing conditions. For convenience the visual field may be divided into: (a) the central field containing the visual task and its immediate background such as the page of a book where the printed words being read are seen against a background area; (b) the surroundings of the task extending out 30 degrees from the visual axis; and (c) the peripheral field beyond the surroundings extending about 70 degrees (for our purpose) out from the visual axis. High brightness contrast between the object and its immediate background is desirable for ease of seeing. Luckiesh and Moss (**19**) cite material which shows that surroundings that are either brighter or darker than the central field decrease visibility of the objects to be discriminated. Apparently it is worse for the surroundings to be brighter than to be darker than the central field. Lythgoe (**23**) also found that dimly lighted or dark surroundings reduced visual acuity in comparison with surroundings of the same brightness as the central field. Luckiesh (**15**) also presents data which show that brightness of surroundings affects accuracy of visually controlled movements. Settings of a visually controlled mechanism showed 12 percent greater error when the surroundings were dark than when the same brightness as the task. The greatest effect was for changes within 15 degrees of the visual axis. He (**15**) also demonstrated that looking alternately at light and dark surfaces where an object must be discriminated leads to rapid reduction in ease of seeing.

All these data lead to the conclusion that vision is best when surrounds are the same brightness as the central field. Surrounds should not be brighter than the central field. Brightness ratios between the central field and the surrounds should not be more than five to one. Ratios greater than ten to one should be avoided. According to Moon and Spencer (**24**), this brightness ratio should not exceed three to one for best vision. It is pointed out that one sees best and visual fatigue is reduced to a minimum when the entire field of view is of approximately the same luminosity as that to which the center of the eye is adapted (i.e., as vision is directed to the work surface).

In parts of the peripheral field outside the surrounds (beyond 30 degrees from the visual axis), the brightness may diminish considerably without materially affecting visual discrimination or ease of seeing, according to Luckiesh (**15**).

Summary

Glare from bright spots in the field of vision not only produces discomfort but also reduces visual efficiency. Brightness contrast between areas within the visual field also reduces visual efficiency and causes discomfort. Illumination should be as uniform as possible and brightness contrast between areas should be reduced to a low minimum. A ratio of not over 3 to 1 between the central visual field and surrounds is best. Light fixtures in the field of vision should have a surface brightness of not over two candles per square inch. It is desirable that it be not over one candle per square inch. In other words, all glare should be eliminated and brightness contrasts kept at a low ratio.

Coordination of Intensity and Distribution

The results of Holway and Jameson (**12**) led them to conclude that four important lighting conditions should be controlled for easy and comfortable visual work. They are glare, brightness contrast, quality of light, and intensity. Intensity, if adequate lighting is to be provided, must always be coordinated with distribution. To increase intensity without controlling distribution will only make a bad situation worse. In fact, when the distribution is poor, relatively low intensities must be employed to avoid visual discomfort (**30**). Furthermore, it is difficult to provide adequate distribution of light with installations above 50 f.c. (**31**).

Color

Introduction

In planning any living or working environment, color as well as lighting must be considered. We are concerned with the following aspects of color: (a) pleasingness, psychological meaning of color, and color combinations; (b) color of paint as a reflecting surface; (c) influence of quality of light on colored objects; (d) coordination of lighting with color of paint.

Pleasingness and Meanings of Colors

Poffenberger (**26**) has summarized the data on pleasingness and meaning of colors. Washburn's study (p. 442) measured the pleasingness of 19 saturated colors, 18 tints, and 18 shades. On the average the tints were most pleasant, the shades next, and the saturated colors least. The size of the colored area had some influence on pleasingness. With saturated colors, the small areas were preferred, but the larger areas of tints and shades were preferred. In saturated colors, red, orange-red, and green-blue were most preferred while yellow and yellow-green tended to be disliked. The most pleasing tints were blue, blue-violet, violet, and red-violet. Tinker (**29**) found that, with the exception of yellow-green, any color was preferred over black, white and gray, i.e., over achromatic surfaces. So-called warm colors are the most frequently preferred. *In all these experiments relatively small areas of color were used.* It is not certain that the same results would hold for areas as large as walls of rooms. There is a suggestion, however, in the fact that tints are more pleasant than saturated colors or shades, and tints are more pleasant in larger areas.

Colors Carry Meaning

Red-orange, orange, and red are considered warm colors, and blue-green, black, gray, and white are cool colors (**29**). In another report (**26**), reds, oranges, and yellows are rated as exciting colors, yellow-green, green, and blue-green as tranquilizing, and blues and the violets as subduing. Apparently these meanings of colors are consistent and compelling.

Color in Reflecting Surfaces

Color as color (hue) in objects and backgrounds has no effect on ease of seeing. Tinker and Paterson (**28**) have shown that when words are printed in colored ink on a colored background, the rate of perceiving the printed material is greater, the greater the brightness contrast between ink and paper irrespective of the colors used. Similarly, the reflection factor of walls, ceilings and furnishings of any living or working space is more important than the color used. These surfaces are virtually secondary sources of illumination. They receive and reflect about the room light from the lamps. Oetting (**25**) points out that eye comfort and performance are more dependent upon the brightness pattern of a working environment than upon the color scheme. This brightness pattern is determined by the intensity and distribution of light from the fixtures and by the reflectance values of all finishes in the room. Oetting (**25**) also points out that all paint should have a flat or mat finish to avoid undesirable specular reflection. Areas above the field of vision should have a surface reflectance of 80 percent or more (white or near white). Walls should have reflectances of 50–60 percent and desk tops of working surfaces, about 35 percent. Tints of appropriate colors may be used. Psychologically, a "change of pace" is desirable in decoration. This can be achieved by varying the color and at the same time avoiding undesirable brightness contrast (see above). There should be a coordination of illumination with colors that are pleasing and have appropriate meanings.

Influence of Illumination on Perceived Color

The color appearance of objects under artificial illumination depends upon the color components in the illuminant and upon the reflectance characteristics of the objects. Most objects do not have the same color appearance under fluorescent and incandescent light as in daylight. Under most fluorescent light, the color of natural objects as food, flowers and human complexions, becomes unnatural and frequently disagreeable. The fluorescent color quality is described psychologically as cold and harsh. Its blue component is detrimental to color tints and tends to make them look bluish or grayish. It takes the warmth and softness out of color. The daylight fluorescent tube has been found to be decidedly unflattering to the human complexion (**2**). Under incandescent light, however, the alterations of color usually are not objectionable. It is kind to tinted wall decorations, tending to enhance soft, warm appearances. It is important to combine color and lighting so the colors in paint do not go "flat" in the light. Buck and Thayer (**4**) point out that one of the more acceptable methods of designating color of a light source is in Kelvin units. For instance, 3500° K means the color from a black body raised to 3500° temperature. The most commonly used illuminants in 1947 are from the white fluorescent lamps, 3500° K, 4500° K, 6500° K (daylight), soft white, and incandescent lamps averaging about 2900° K. In a later study, Buck and Froelich (**3**), after determining the color characteristics of the human complexion, obtained preferences for lights while viewing a model with "average" complexion. The percent distribution of choices were 45 for soft white fluorescent, 24 for incandescent, 15 for 4500° K, 14 for 3500° K and 2 for 6500° K (daylight). The light chosen most frequently, soft white, had a spectral character close to that of the average complexion. Even this had much more blue and less red than the complexion. These authors also found that lamps preferred for viewing colored materials were in general very nearly the same dominant wave-lengths as in the average complexion.

In general, a light which enhances warmth and softness of colored paints is desirable. It should be one that does not change natural colors (including complexion) in disagreeable ways.

Coordination of Lighting With Color of Paint

To achieve a pleasant environment and at the same time to maintain good seeing conditions, it is necessary to coordinate lighting with decoration. According to Kahler and Meacham (**13**), good seeing conditions are attained when there is adequate light on the task and when brightness of the task is no greater than three times that of surrounds (30 degrees each way from line of vision). The ratio of the brightness of a light fixture to its background also should not be greater than 3 to 1. These authors (**13**) point out that there is no uniform agreement that best vision occurs when the brightness ratio is one to one for the task and the surrounds. Many consider that three to one is as satisfactory as one to one. In decoration, some contrast is desirable. Lack of variation in brightness in decoration tends to be monotonous and uninteresting. Contrast is one of the important tools of the decorator. The blending of high-lights and shadows adds attractiveness to the living space. Kahler and Meacham (**13**) carried out studies to determine whether maximum ratios of three to one between task and surrounds or between adjoining areas in the visual field are practical from a lighting standpoint or desirable from the decorator's viewpoint. Findings indicated that, in general, good decorative schemes can be executed with low brightness ratios; but this is not so with a one to one ratio, because some contrast is necessary to give character and interest to a room. Ratios can be kept less than three to one or only slightly greater.

Applications to Submarines

In considering application of data on illumination and color to a living or working place, it is necessary to take into account the specific characteristics of each situation. In many ways the submarine is a unique situation. The ceilings tend to be low, much of the ceiling and wall space is occupied by pipes, conduits, dials or other machinery. The deck space is confined largely to narrow passageways and walks. Consequently reflecting surfaces are at a minimum. Also, it is difficult to find adequate space for light fixtures. In view of these facts, one cannot make ready applications of data worked out for homes, offices, and factories. Special problems must be met.

In this part of the report, it is not planned to specify in detail the lighting and color of paint in submarines, but the illumination and color requirements will be pointed out. The physical means for providing the light and paint must be taken care of by specialists.

A survey of the lighting and painting in several submarines revealed rather uniformly poor conditions. The most common deficiencies were found to be: insufficient light at the places where visual discriminations have to be made; marked brightness contrasts in adjoining surfaces; lack of adequate uniformity in distribution of illumination; disturbing specular reflections from shiny surfaces such as highly polished valve and switch handles; and glare from excessively bright and poorly located light sources. It is not uncommon to find a bare 50- to 75-watt incandescent bulb hanging just above the line of vision. The tendency of the men in such a place as the control room to prefer relatively very dim light is undoubtedly largely due to uncomfortable specular reflection (glare) from highly polished metal parts when the illumination is brighter. All these unsatisfactory conditions result in visual discomfort which may accentuate nervous tension. Decorations tend to be dark, dingy and psychologically forbidding. The pea-green of newly painted jobs is a repulsive color to most people. Apparently little thought has been devoted to providing a cheerful and pleasing paint job.

General Considerations

Ordinary light fixtures are not suited in general for installation in submarines because of low ceilings and restricted space (between pipes, etc.). It is necessary, therefore, to design some new fixtures if adequate lighting is to be achieved.

One important question that must be raised is whether fluorescent lights should be used. After exposure to direct fluorescent light, recovery of sensitivity in the night vision apparatus is retarded. After 30 minutes in the dark, night vision is only half as good as when previously exposed to incandescent light. This fact is very important in situations where rapid adaptation to night vision is desirable, i.e., use of periscope at night, look-out duty, etc. Secondly, most people do not like the harsh, cold appearance of fluorescent light. Human complexion and colored objects look unnatural under many fluorescent tubes (3500° K, 4500° K, 6500° K—daylight). It must be recognized, however, that Buck and Froelich (**3**) did find that people preferred soft white fluorescent light over incandescent. On the other hand, the heat generated and their shape seem to be the main arguments against incandescent fixtures. Air conditioning should take care of the heat. The quality of incandescent illumination is usually well liked. It may be described as soft and warm, qualities that are comfortable and relaxing to humans. Careful thought and some research should be done on this problem before fluorescent lighting is generally adopted. It should be noted that incandescent light fixtures are now available in shapes other than the standard bulb.

In those compartments where the ceilings are low or predominantly covered by pipes and the like, there should be no attempt to illuminate the ceiling to obtain reflected light, and in addition a low bright ceiling makes an uncomfortable visual environment. Furthermore, the brightest illumination should be at the work surface, not on a low ceiling which should be the least conspicuous place in the room.

Uniform general lighting in all compartments of submarines is impracticable and probably impossible. While best vision results from brightness ratios of three to one or less for adjoining areas in the visual field, ratios as great as ten to one are tolerable. And as pointed out above (**13**), some contrast in color and brightness is desirable. Nicely blended shadows increase the attractiveness of a living space.

In much of the submarine, directional lighting will be found most satisfactory. Any fixtures within the field of view should have diffusing surfaces or louvers so that the surface brightness will be low, i.e., one candle per square inch, or as near this as is possiple. What is needed is directional lighting produced by sources of low brightness at normal viewing angles, and of sufficiently low brightness in the "beam" so that specular reflection will not become a problem. The important thing is to avoid glare from fixtures and to get the right amount of light to places where it is needed, i.e., dial-faces, table and desk tops, etc. Where directional lighting is used, care must be exercised to avoid excessive brightness contrast with adjoining areas. When directional light is applied to glass covered dial faces, the angle of incidence should be such that the greatest amount of light possible gets through the glass to the dial surface and still have the angle of reflection such that the reflected light does not shine into the reader's eyes. If practical, it would be desirable to dispense with highly polished surfaces (such as valve handles) to avoid annoying specular reflection. They might be painted, or a dull-appearing metal might be used.

The amount of light needed will vary with the fineness of the discrimination required. Where scale divisions on dial faces are relatively small, where reading of small print or maps is required, where typing is to be done, and where labels in small print are to be perceived, the light should be between 20 and 50 foot-candles. Where the discrimination is less severe, 10 to 20 foot-candles will be found adequate. The illumination on any work surface should not drop below 10 foot-candles unless requirements of the situation demand it, e.g., in the conning tower.

Painting should provide both reflecting surfaces (where possible) and pleasing color combinations. On large areas saturated colors and any paint of low reflectance should be avoided. Tints of colors (pastels) should be used. Ceilings that are exposed so they may act as reflecting surfaces should be near white (perhaps a light cream or ivory) with a reflecting factor of 70 to 80 percent. The paint on instrument panels and bulkheads should reflect 40 to 60 percent, and on table tops and other furniture it should reflect 30 to 40 percent of incident light. If there are any sizable deck spaces, the reflection factor should be around 20 to 25 percent. To reduce apparent specular reflection, instrument dials should be white with black markings, because the glare image on cover glass will be less annoying and in many cases the visibility of scale markings will be greater. Instrument panels should be light enough (gray or other color) to avoid sharp contrast with the white dial face.

Variety in color of decoration is desirable. To maintain a pleasing environment and to get pleasing contrasts in decoration, different colors should be used in the same compartment, and still different combinations in other compartments. Tints and light grays possess several advantages for decorating submarine compartments: tints (light colors) make a room appear more spacious (**13**), and the ceilings appear higher. Only flat (soft or mat surface) paints should be used to avoid annoying specular reflections. Some of the favored color tints are buff with a little umber, ivory, cream, bluish. A good combination is coral and peach. Unfortunately, there are practically no experimental data on pleasingness of tints in large areas (as on a wall) or of combinations of tints. A light gray will probably give a fairly pleasing chromatic contrast with several tints.

Light tints, such as ivory, tend to darken after a few months in the submarine. Paint laboratories should be given the task to determine paints which will keep their brightness longer when used in submarines.

Specific Applications

In the above section, consideration was given to principles that can find application in various parts of a submarine. In this section, attention will be given to additional factors that must be taken into account in specific compartments.

The Conning Tower

The conning tower is a rather special situation. While the light on vital instruments must be enough for accurate and rapid vision, the eyes of the periscope observer must be adapted for accurate seeing in the light coming through the periscope, i.e., be adapted to a brightness equal to or slightly less than that outside. At the same time, there should be enough ambient illumination to facilitate moving about safely in the conning tower. It is best that no direct light reach anyone's eyes. Light on vital instruments can be controlled direct rays or internal illumination.

There are indications that illumination need not be as dim as has previously been thought. The study of Verplanck (**32**) on use of red illumination in the submarine conning tower is pertinent. Speed and accuracy of locating a target with a periscope were measured when the conning tower (experimental set-up) was illuminated with two 50-watt red steam-tight fixtures and all instrument lights were on. Illumination varied from .02 to .50 f.c. from one location to another. The target was in a dimly illuminated skyscape. Observers were just as accurate and just as fast when adapted to the red light as when adapted to darkness. It was concluded that no untoward effects on night vision occur when there is this amount of red light (two fixtures, plus illuminated instruments) present. This was true whether the interior of the conning tower was painted black or white. It is pointed out in a footnote, however, that this amount of light would not be used in a night approach or when a submarine is surfaced at night.

Fluorescent fixtures that are not shielded from workers' eyes should not be employed in the conning tower. The paint here can be fairly bright, i.e., light grays and pastel colors (see above).

Maneuvering Room

There may be some overhead light fixtures for general illumination. These fixtures should be such that the surfaces presented to the eye have low brightness values. General illumination of the compartment need not be more than 5 to 8 f.c. The gages and any other working surfaces should be illuminated by additional directional light so arranged as to avoid glare. There should be not less than 20–30 f.c. of light on any working surface where visual discrimination is required. As noted above, dial faces should be white and instrument panels light (probably gray). See "General Considerations" above for color and brightness of paint.

Control Room

Suggestions for maneuvering room apply here.

Wardroom

The situation is a little difficult here, since the ceiling is low and general illumination of a fair degree of brightness is needed. There should be 20 to 30 f.c. at the table surface. Fixtures should be devised that have low surface brightness, that spread some light sidewise, and that yield adequate direct light to the table surface. Quality of the light should be such that the human complexion is seen in its natural color. The paint should be cheerful and warm tinted, with some hue contrast to make the room a pleasing environment.

Crew's Mess

There should be a fairly bright light at the table top, since the men read as well as eat here. The surface of the fixture yielding general illumination should be of low brightness and the quality of the light should be such that food and human complexions look normal in color. Supplementary lighting may be needed to provide the desired 20 to 25 f.c. on the table top. In this room, also, the paint should be light tints that appear warm with some contrast in hue from area to area.

Radio Room

The general illumination should be well diffused and of perhaps 8 or 10 f.c. There should be local directional lighting of 30 to 35 f.c. at the typewriter. Pleasing colors should be used on the little space available.

Bunk Room

Considerable effort should be exerted to made this a more pleasing living space. There should be well-diffused general illumination from fixtures that do not provide glare spots. In addition, there should be individual bunk light fixtures that are shielded (from vision of others) and with directional adjustment. Possibly it would not be feasible to have more than 10 to 15 f.c. of light at the reading surface. This will be adequate for reading the type in books and magazines. Pleasing combinations of pastel colors should be used for painting.

Skipper's Cabin

Only directional lights should be employed, yielding an illumination of 25 f.c. at the desk and about 15 f.c. from reading lamp at the bunk. The latter fixture should be an adjustable directional light. Paint should be in warm appearing tints.

Engine Room

This room permits adequate fixtures for well-diffused general illumination. Surface brightness of these fixtures must be kept low. General illumination should be about 15 f.c. In addition, there should be local directional lighting where the more exacting visual discriminations must be made, i.e., dial faces, etc. Painting should be pleasing and light enough for adequate reflecting surfaces (see above).

Torpedo Rooms

Fixtures of low surface brightness should be used for general diffused illumination of around 10 to 15 f.c. In addition, there should be local directional light where finer visual discrimination is required. Paint should furnish good reflecting surfaces. Use some color for pleasing contrasts.

Head

Shield the light fixture so it will not be a severe glare spot such as is ordinarily found there.

General Summary

The spectral quality or color of light in ordinary illuminants is relatively unimportant. Only for threshold seeing is yellow slightly better than other artificial illuminants. In ordinary seeing situations which are supraliminal as in reading, one light is as good as any other: tungsten filament, fluorescent, mercury vapor, daylight. There is, however, a psychological objection to most fluorescent light because of its "harsh" and "cold" appearance.

Intensity of illumination has been considered in relation to visual acuity, size of object to be discriminated, speed of vision, brightness contrast, and efficiency of performance. The trend of the data reveal that visual efficiency of normal eyes increases rapidly as illumination is increased from a fraction of one up to 5 f.c., more slowly from there up to about 10 f.c., and very slowly from 10 to about 20 f.c. Thereafter, the improvement in vision is slight when the object to be discriminated is of moderate size (3 to 6 minutes). When the object is smaller, however, visual efficiency improves by practical amounts up to 40 or 50 f.c. And the greater the brightness contrast between object and background, the better the visual efficiency. Certain visual functions like visual acuity and speed of vision improve by amounts that are insignificant in practical situations up to more than 1000 f.c. for the former and more than 100 f.c. for the latter.

Visual efficiency and visual comfort are reduced by two factors of distribution: glare from bright spots in the field of vision, and marked brightness contrast between areas in the field of vision. Illumination should be as uniform as possible, and brightness contrasts between adjoining areas should be small. A ratio of not more than three to one between task and surrounds is best. To avoid glare, the surface brightness of light fixtures should be as close to one candle per square inch as possible. There should always be a coordination of intensity with distribution in planning illumination.

Painted walls, ceilings and furnishings should be pleasing and should furnish reflecting surfaces. Certain colors are preferred over others, tints are preferred to saturated colors, and tints are preferred in large rather than small areas. The paint should be light enough to reflect a large portion of incident light: ceilings 70 to 80 percent, bulkheads 40 to 50 percent, desk tops 30 to 40 percent and floors 20 to 25 percent. Psychological variety can be achieved by using contrasting colors in adjoining areas and from room to room.

A light that enhances warmth and softness of colored objects is desirable. Also, the illumination should be one that does not alter markedly the color of natural objects as complexions and foods. Good decorative schemes can be executed with a low brightness contrast of around three to one.

The research findings cited above find many applications in the lighting and painting of submarines. A number of these are listed.

Problems for Research

There are a number of problems in lighting and color in submarines that need further research. Those which seem of greater importance to the writer are listed below:

1. The pleasingness of colors and color combinations has been studied in the psychological laboratory, employing relatively small areas of colored stimuli. No controlled experimentation has been carried out to discover the pleasingness of large areas (walls of rooms) of color and color combinations. This should be done. The present use of colors is largely in terms of opinions of artists and decorators.

2. Light tints of colors and white paint rapidly turn much darker in the submarine. A paint laboratory should be given the task

of finding paints which do not darken rapidly on exposure to gases in submarines.

3. It is reported that the recovery of night vision after exposure to direct fluorescent light is slow. Further experimental check of this relationship should be made before fluorescent light is adopted for use in submarines and other warships.

4. Much experimentation must be done with newly devised light fixtures for different parts of the submarine. These should include fixtures for general lighting in compartments such as the wardroom and crew's mess; directional fixtures for illumination of dials and other important areas; and directional bunk lights for both officers' staterooms and crew's bunkroom.

References

1. Arnold, D. C., & Tinker, M. A. The fixation pause of the eyes. *J. exp. Psychol.*, 1939, **25**, 271–280.
2. Birren, F. The specifications of illumination and color in industry. *Trans. Amer. Acad. Ophthal. Otolaryng.*, 1947, **51**, 242–251.
3. Buck, G. B., & Froelich, H. C. Color characteristics of human complexions. Illum. Engng., 1948, **43**, 27–47.
4. Buck, G. B., & Thayer, R. N. Color technology of fluorescent lamps. *Illum. Engng.*, 1947, **42**, 415–434.
5. Cobb, P. W. Some experiments on speed of vision. *Trans. illum. Engng. Soc.*, 1924, **19**, 150–175.
6. Cobb, P. W., & Moss, F. K. Glare and the four fundamental factors in vision. *Trans. Illum. Engng. Soc.*, 1928, **23**, 1104–1120.
7. Ferree, C. E., & Rand, G. The power of the eye to sustain clear seeing under different conditions of lighting. *J. educ. Psychol.*, 1917, **8**, 451–468.
8. Ferree, C. E., & Rand, G. The effects of variations of illumination on functions of importance to the working eye. *Trans. Illum. Engng. Soc.*, 1920, **15**, 769–801.
9. Ferree, C. E., & Rand, G. Intensity of light and speed of vision studied with special reference to industrial situations. Part I. *Trans. Illum. Engng. Soc.*, 1927, **22**, 79–110.
10. Ferree, C. E., & Rand, G. Visibility of objects as affected by color and composition of light. Part I: With lights of equal luminosity or brightness. *Person. J.*, 1931, **9**, 475–492; Part II: With lights equalized in both brightness and saturation. *Person. J.*, 1931, **10**, 108–124.
11. Ferree, C. E., Rand, G., & Lewis, E. F. The effect of increase of intensity of light on the visual acuity of presbyopic and non-presbyopic eyes. *Trans. Illum. Engng. Soc.*, 1934, **29**, 296–313.
12. Holway, A. H., & Jameson, D. *Good lighting for people at work in reading rooms and offices.* Boston (Soldiers Field): Harvard Business School, Division of Research, 1947.
13. Kahler, W. H., & Meacham, J. A. Correlation of brightness ratios and decoration. *Illum. Engng.*, 1948, **43**, 175–193.
14. Luckiesh, M. *Light, vision and seeing.* New York: D. Van Nostrand Co., 1944.
15. Luckiesh, M. Brightness engineering. *Illum. Engng.*, 1944, **39**, 75–92.
16. Luckiesh, M., & Moss, F. K. Seeing in sodium-vapor light. *J. opt. Soc. Amer.*, 1934, **24**, 5–13.
17. Luckiesh, M., & Moss, F. K. Seeing in tungsten, mercury and sodium lights. *Trans. Illum. Engng. Soc.*, 1936, **31**, 655–674.
18. Luckiesh, M., & Moss, F. K. *The science of seeing.* New York: D. Van Nostrand Co., 1937.
19. Luckiesh, M., & Moss, F. K. Brightness-contrasts in seeing. *Trans. Illum. Engng. Soc.*, 1939, **34**, 571–597.
20. Luckiesh, M., & Moss, F. K. Vision and seeing under light from fluorescent lamps. *Illum. Engng.*, 1942, **37**, 81–88.
21. Luckiesh, M., & Taylor, A. H. Radiant energy from fluorescent lamps. *Illum. Engng.*, 1945, **40**, 77–88.
22. Lythgoe, R. J. *Illumination and visual capacities.* Special Report No. 104, Medical Research Council. London: H. M. Stationery Office, 1926.
23. Lythgoe, R. J. *The measurement of visual acuity.* Special Report No. 173, Medical Research Council. London: H. M. Stationery Office, 1932.
24. Moon, P., & Spencer, D. E. The visual effects of non-uniform surrounds. *J. opt. Soc. Amer.*, 1945, **35**, 233–248.
25. Oetting, R. L. Light finishes improve office performances. *Office Management and Equipment*, 1947, October, 61–66; 106.
26. Poffenberger, A. T. *Psychology in advertising.* New York: McGraw-Hill, 1932.
27. Simonson, E., & Brozek, J. Effects of illumination level on visual performance and fatigue. *J. opt. Soc. Amer.*, 1948, **38**, 384–397.

28. Tinker, M. A., & Paterson, D. G. Studies of typographical factors influencing speed of reading. VII. Variations in color of print and background. *J. appl. Psychol.*, 1931, **15**, 471–479.

29. Tinker, M. A. Effect of stimulus-texture upon apparent warmth and affective value of colors. *Amer. J. Psychol.*, 1938, **51**, 532–535.

30. Tinker, M. A. Illumination standards for effective and comfortable vision. *J. consult. Psychol.*, 1939, **3**, 11–19.

31. Tinker, M. A. Illumination standards. *Amer. J. publ. Hlth.*, 1946, **36**, 963–973.

32. Verplanck, W. S. An investigation of the red illumination of the submarine conning tower. Interval report No. 1, Bureau of Medicine and Surgery, Project X-519 (Av-273-P), 1946.

33. Weston, H. C. *The relation between illumination and industrial efficiency. 1. The effect of size of work.* Medical Research Council Report. London: H. M. Stationery Office, 1935.

34. Weston, H. C. *The relation between illumination and visual efficiency—The effect of brightness contrast.* Industrial Health Research Board Report, No. 87. London: H. M. Stationery Office, 1945.

Chapter 18

MOTION SICKNESS: ITS PSYCHOPHYSIOLOGICAL ASPECTS

JAMES E. BIRREN

U. S. Public Health Service

Introduction

Most people with sea experience can describe the symptoms of seasickness, but it is less well known that the same symptoms may be produced in other situations involving acceleration. The motion of airplanes, streetcars, automobiles, swings and other sources of movement may cause sweating, nausea and finally vomiting. Marked lethargy and even headache and dizziness may also occur in some persons. The whole pattern of symptoms may be called *motion sickness,* whether produced in an airplane or aboard ship.

Statistically there is nothing unusual about motion sickness, since more than half of the population may be made seasick and some investigators believe that everyone may be made motion-sick under appropriate conditions. Motion sickness is therefore a common psychophysiological phenomenon. Animals, as well as man, share this predisposition to illness when exposed to periodic motion. As in man, there are individual differences in susceptibility. Some dogs become motion-sick after a few minutes in a simple laboratory swing, whereas others will not salivate and vomit even after a half hour of continuous swinging.

While the primary factor in motion sickness is acceleration, there are also secondary factors which may influence the incidence and severity of motion sickness. Although it is unlikely that these secondary factors, such as vision, odors, and the attitudes of the individual, can by themselves produce motion sickness, it is recognized that they may well influence its course given the primary conditions of acceleration.

Primary Sense Organ in Motion Sickness

The vestibular apparatus of the inner ear may be regarded as the primary source of sensory stimulation in motion sickness. The evidence is as follows: (1) deaf mutes are immune to motion sickness, (2) animals may be immunized by removal of the vestibular apparatus or by severing the eighth cranial nerve, (3) removal of the vestibular portion of the cerebellum eliminates motion sickness in dogs, and (4) head position affects the incidence of motion sickness (**19, 55, 82, 102, 114**).

Particular sensory receptors of the inner ear that are involved in motion sickness are not yet clearly defined. The inner ear is a relatively complicated structure and the sensory portions consist of the cochlea, saccule, utricle and semicircular canals (**26**); of these only the utricle and the semicircular canals are generally considered as being important in the production of motion sickness. Each sensory unit of these receptor organs consists of a mass, gelatinous or calcareous, attached to the hairs of a sensory epithelium. When the body is moved, the suspension medium permits the "mass" to exert a stimulating pressure upon the sensory endings. Experiments on nerve fibers from both the semicircular canals and the utricle have demonstrated that these organs are excited by acceleration, and in the case of the utricle, by gravity as well. The sensory endings in the semicircular canals appear to be particularly affected by angular acceleration, whereas the utricle also is excited by linear acceleration and gravity. Either angular or linear acceleration may lead to motion sickness, and in most situations aboard ship it is likely that the stimulus is a combination of these two forms. (See Part II of this chapter for additional details.)

A close relation exists between the vestibular apparatus and the autonomic nervous system. Various phenomena associated with the autonomic nervous system, other than motion sickness, e.g., vasomotor changes and pupil dilation, may be seen in animals upon stimulation of the vestibular apparatus electrically, by hot or cold water, or by acceleration (**105**). The role of the higher brain centers in vestibularly induced responses appears to be primarily an inhibitory one (**120**).

Effects on Efficiency

The undesirable effects of motion sickness upon efficiency of personnel have often been dismissed as negligible, since few Navy men fail to stand their watches during emergencies because of seasickness. The exception to this is the group of men who are seasick almost all the time at sea (**100, 101**). Even one such individual may represent a serious loss to a submarine because of the specialized skills and limited number of personnel aboard. For this reason, captains of submarines on War Patrols felt that men sent to boats should be carefully screened to eliminate any men who might be chronically seasick (**30**).

There is some evidence that most individuals who are made transiently motion sick can exert themselves to the point of adequate performance when the situation demands (**5**). Such performance, which might be called "peak efficiency," need not be closely related to the performance of daily routine. The latter might be called "maintenance efficiency." Even though seasick, a man may be able to exert himself to emergency performance; the result of his efforts depends to a significant extent upon how well he has maintained his "gear." During rough weather, seasick personnel lose interest in doing anything except the barest necessities, and an obvious lack of spontaneity can be observed aboard ship even in those men not frankly seasick. Not only do the men fail to indulge in the usual "horse play" and spend almost all time off-watch in their bunks, but they also fail to secure gear properly. Such effects of seasickness upon "maintenance efficiency" should be seriously considered as having a significant effect upon personnel efficiency or performance.

Criterion of Motion Sickness

Every study of motion sickness has been confronted with the issue of deciding when a man is motion sick. It is obvious that different results will be reported on the incidence of motion sickness, or the beneficial effect of some therapy, if a symptom such as sweating is used as the criterion instead of employing vomiting as the criterion of motion sickness. The following is a general outline of useful schemes to rate degree of motion sickness:

Motion Sickness Rating	
0	No symptoms
1	Slight nausea or other minor complaint
2	Nausea and sweating
3	Vomiting but able to work
4	Incapacitated

Only gross sweating, pallor, and vomiting are observable without questioning the individual. Sweating may be detected by means of skin resistance measurement employing a galvanometer (**41**). This technique, however, is limited largely to laboratory conditions. The association of sweating with motion sickness is much more regular than are changes in blood pressure and pulse rate (**40, 41**). Measurement of blood pressure and pulse rate were reported to be of little or no value in either estimating severity of motion sickness or predicting susceptible AAF personnel during exposure to motion (**40**). A characteristic fall in blood pressure was reported on a small group of seasick men (**104**), but it is unlikely that such changes are useful in estimating degree of motion sickness in view of Hemingway's findings (**40**).

Sources of Information

The literature on motion sickness published prior to 1942 was summarized by McEachern (**77**) and McNally (**82**). A more recent, comprehensive, review has been prepared by Tyler and Bard (**114**). Aside from research carried on by the military services, the Subcommittee on Motion Sickness of the National Research Council sponsored research projects and published a bibliography on this topic (**92**). A similar subcommittee was organized in Canada. Both of these groups have published research reports and minutes of their meetings which are filed with the National Research Council in Washington, D. C. Several specialized reviews of service-conducted research have been made (**2, 48, 79, 85**). In addition to the bibliographies of the reviews just cited, a comprehensive list of titles is contained in the bibliography on aviation medicine (**52, 53**).

Importance of Motion Sickness

The importance of motion sickness may be judged from three types of statistics: (1) the number of men disqualified for service because of severe and persistent motion sickness, (2) the incidence or number of personnel motion sick on one or on a group of operations, and (3) the distribution of susceptibility to motion sickness in the population. The latter distribution is in a sense the percentage of persons who would ever be made motion sick.

Statistics on the incidence of motion sickness show wide variations; although this is not necessarily the result of unreliable observations. Such factors as adaptation, type of craft, and weather conditions all have an important effect upon incidence.

Disqualification for Military Service

Seasickness

Although a considerable number of men were transferred to shore duty by the United States Navy during World War II because of intractable seasickness, no records were kept of the men so transferred during the early part of the war. Evidence that the number of men was significant is contained in the fact that several of the naval hospitals in the larger ports instituted special procedures for examining men sent there because of severe and persistent seasickness.

A summary of Submarine War Patrol Reports mentions seven men who were transferred ashore from different submarines (**30**). The circumstances under which these men were transferred suggest that the men were of little or no value to the boat. Of 1,471 submarine patrol reports summarized by Duff, 35 reported seasickness as a problem.

In a discussion of motion sickness prepared for the Royal Australian Air Force, McIntyre cites no data for disqualification due to seasickness, although he viewed the problem as of "considerable military significance" (**79**). This lack of specific information is less true of the air services because of the greater opportunity to observe each man during the training of flying personnel.

Airsickness

Available data suggest that motion sickness was a major cause of disqualification in the air forces of the Allied Nations during World War II. The "washout" or "scrub" rate for airsickness in pilots and air crew men has been variously reported from 0.15 to 6.5 percent (**79**). Of all types of disability in 2,399 cases analyzed in the South African Air Forces, ". . . airsickness is the most important single disability . . . with the highest number of permanent limitations in the air crews" (**95**). Of 2,682 pilot trainees in the RCAF, 0.5 percent were eliminated because of airsickness. Navigators appear to suffer a higher rate of disqualification; about five percent of the potential navigators are disqualified because of airsickness according to United States and Australian reports (**74, 79**), and air gunners were reported to have the highest rate of disqualification for airsickness (**59**). A more recent report on the subject indicates that flying personnel in general have a two percent rate of disqualification because of airsickness (**49**). About six percent of naval pilot trainees were reported to have sufficient airsickness during training to constitute a problem (**118**). While this summary is not comprehensive nor does it establish rates to be expected for latest types of plane design, it does indicate clearly that motion sickness is a significant problem in the air services.

Sea- and Air-Borne Troops

Air- and sea-borne troops are a special case of the problem of motion sickness since these personnel are for the most part land based until they are transported to areas where they may need to engage in immediate action. They thus lack whatever adaptation to motion sickness may be acquired by regular exposure to the motion of the transporting craft. While there are no data on the number of men disqualified for this type of duty, an estimate might be made from the number of men incapacitated during any single operation. During one airborne infantry maneuver in mild weather, 9.8 percent were airsick; of these three percent were unable to carry out their duties (**72**).

Incidence of Motion Sickness

The incidence of motion sickness varies considerably from day to day due to variation in the roughness of the sea or air. To be definitive about the incidence of motion sickness, therefore, it would be necessary to describe in detail the circumstances under which the particular incidence was observed. In the case of seasickness, it would be pertinent to have such information as: the condition of the seaway, the size and type of ship, position and activity of personnel while underway, and selection of personnel.

A large factor in determining the magnitude of a reported incidence of motion sickness is the original informant. The incidence is higher when the individuals themselves are the informants than when the ranking non-commissioned or petty officer reports. It is still lower if the division (Navy) or company (Army) officer reports and lowest if the medical officer reports. The medical officer tallies only those cases who report to the medical department for treatment, in general, those men most seriously affected. Since, in the past, there were no effective remedies available to the medical officer, it was common knowledge among military personnel that going to the medical department did little good. Hence, usually only new personnel went to the medical officer for treatment of motion sickness.

Seasickness

Seasickness on landing craft (LCVPS) engaged in amphibious training was reported by Tyler (**111**) to be high even in relatively smooth seas. His large-scale studies involved over 15,000 men. An overall incidence of 35 percent seasickness in untreated personnel was noted; 13 percent were severely seasick, i.e., had severe nausea with or without vomiting. On individual test runs the incidence of seasickness went as high as 53 percent.

An incidence of 16 percent seasickness in troops on assault craft was observed in calm seas (**51**), and rose as high as 51 percent in soldiers taken to sea in rough weather (**1**). In the latter case, 45 percent of the men observed vomited.

Airsickness

Glider personnel apparently suffer a high incidence of airsickness. On different flights it ranged from 30–70 percent among RAAF glider personnel, as opposed to 15–30 per cent of flight training personnel in the RCAF (**79**).

On glider trips of less than one hour, an incidence of 30–50 percent has been reported "moderately bumpy" trips (**67**).

Airborne troops in the United States had an observed incidence of 27 percent on long flights (3½ hours) and 12 percent on short flights (40 minutes) (**43**).

Susceptibility to Motion Sickness

The number of persons motion sick on any single cruise or flight is not necessarily equal to the number of persons who may be motion sick at some time in their life. Thus, the question, "Have you ever been sick?" will be answered "yes" by a higher percentage of naval personnel than might ever be observed to be sick on a single day aboard ship.

A group of 511 naval personnel with sea experience gave questionnaire reports about their frequency of seasickness (Table I) (**19**). Only 192, or 37.6 percent, indicated they were never seasick. Since the average amount of sea duty was nearly two years in this group, the men had had ample opportunity for exposure to all kinds of weather at sea. The seven percent of the personnel who said they were frequently or almost always seasick to the point of nausea and vomiting, represent the population who would be most profitably screened before sea duty.

It is worth while to note that approximately one out of three men will never be seasick. This relatively small number of men who will never be seasick suggests that it would be difficult to screen men to secure a totally seasick-resistant group.

A group of 1,560 male naval personnel were given a questionnaire about their experience with various forms of motion sickness (**19**). Of all types of situations leading to motion sickness, swings were most frequently checked as producing symptoms (Table II). (For any one question, the number of cases is not equal to 1,560, because of the lack of experience and an occasional reject for an uninterpretable answer.) A question about symptoms produced by ice and roller skating was answered by only one individual as almost always making him sick; thus, the items in the questionnaire were presumably answered with high discrimination. The information gathered by this method approximates experimental observations. Of this group, 24.4 percent indicated they had experienced some symptoms on a swing; whereas swing tests on 500 air force personnel revealed an incidence of swing sickness of 28.6 percent (**47**).

Elimination or Reduction of Motion Sickness

There are five general approaches to reducing the incidence of motion sickness: (1) elimination of susceptible personnel through screening procedures, (2) acclimatization of personnel by increasing daily exposure to some form of adapting motion such as that of a swing, (3) use of drugs, (4) design of the craft or vehicle causing motion sickness, and (5) control of auxiliary factors such as position, posture, and vision. Each of these approaches has its special advantages and disadvantages.

TABLE I

Susceptibility to Seasickness in Experienced Naval Personnel

Frequency of Seasickness	*N*	%
Never	192	37.6
Rarely	168	32.9
Occasionally	74	14.5
Often	41	8.0
Almost always	36	7.0
	511	100.0

1. Screening procedures have the advantage of supplying personnel who may require little further attention. To the extent that such procedures are effective, the use of drugs, acclimatization periods, etc., is not required. The disadvantages of screening are that often it is not practical to eliminate men who have special skills, and there may not be a sufficiently large body of men to draw upon to permit a necessary high rate of elimination.

2. Acclimatization to motion is a potentially useful technique in the treatment of

individuals as a preparation for some critical activity when motion sickness might be encountered, e.g., paratroopers. Its limitations lie in the temporary nature of the protection and in the cost of exposing personnel to daily adapting motion wherein it may be necessary to observe each person carefully to avoid over-exposure and thus precipitate motion sickness. Furthermore, the adaptation may be related to the form of motion in a rather specific way so that it is not possible to condition a person's resistance in general.

TABLE II
SUSCEPTIBILITY TO MOTION SICKNESS AMONG MALE NAVAL PERSONNEL

		Never Sick	Sick one or more times	Frequently, or almost always sick (nausea and/or vomiting)
Airplanes	*N*	615	111	20
	%	84.7	15.3	2.8
Automobiles	*N*	1384	172	14
	%	88.9	11.1	0.9
Busses	*N*	1343	197	22
	%	87.2	12.8	1.4
Elevators	*N*	1368	150	29
	%	90.1	9.9	1.9
Ferris Wheels	*N*	1213	233	66
	%	83.9	16.1	4.6
Merry-Go-Rounds	*N*	1191	286	75
	%	80.6	19.4	5.1
Roller Coasters	*N*	988	203	53
	%	83.0	17.0	4.4
Somersaults	*N*	977	230	42
	%	80.9	19.1	3.5
Street Cars	*N*	1259	139	28
	%	90.1	9.9	2.0
Subways	*N*	839	81	12
	%	91.2	8.8	1.3
Swings	*N*	1135	367	107
	%	75.6	14.4	7.1
Trains	*N*	1337	209	15
	%	86.5	13.5	1.0

3. Drugs have the advantage of being administered with little difficulty when the occasion requires the protection of personnel whose skills are valuable or for the comfort of passengers. They lack flexibility for general use among regular personnel, since in most cases drugs must be given before exposure to motion. This is not an objection for passengers, since they may be given the drug before embarkation. In the case of a crew, bad weather would have to be anticipated before the administration of drugs would be justified. Furthermore, the responsibility for the management of drugs must rest with medical personnel who are not always available on small craft, or for that matter, cannot be standing by merely for the possible dispensing of a drug. There are also limitations of dosage, so that repeated administration of the same drug may not be possible in continued rough weather.

4. Design of ships or planes to eliminate motion sickness has not as yet been exploited, since design has been more dependent upon such matters as speed and function. While, in general, there is an inverse relation between size of the craft and the incidence of motion sickness, the use of small planes and boats is a necessity in many instances, e.g., fighter planes and landing craft.

5. Control of auxiliary factors may be used to reduce the incidence of motion sickness when selection and other forms of control are not applicable. Lying in the supine position, viewing the horizon, and placement at the center of gravity of a vehicle, all have their function in reducing motion sickness. It has also been suggested that an artificial pilot be used to orient personnel working below deck (**113**). The disadvantages of such control lie in the fact that not all the crew or passengers can be at the center of gravity nor can personnel be permitted to lie in bunks or view the horizon at will.

Screening Procedures

Various screening procedures have been suggested for detecting men highly susceptible to motion sickness: (1) mechanical devices which attempt to determine the resistance of the individual to motion sickness, e.g., swings, vertical accelerators, and complex roll-rockers, (2) interview techniques employing professional psychiatric and psychological personnel, (3) motion sickness questionnaires that are answered by each man, e.g., during personnel processing upon entrance to military service, (4) observation during exposure to the actual situation involving motion as aboard a ship or airplane, (5) labyrinthine tests, and (6) emetic drugs.

The success of these procedures depends upon the degree of correlation between the various forms of motion sickness. If only a single factor were involved in motion sickness, i.e., periodic acceleration, and individuals had relatively stable thresholds for the tolerance of such acceleration, it would be a simple matter to predict how many, and also which, individuals would be sick in the presence of known accelerations. If the reasonable assumption is made that there is a high correlation among the various forms of motion sickness, there are still several factors which mask the "true relation": (1) the amount of experience with different forms of transportation varies markedly, (2) ships, airplanes, and land vehicles vary in size and speed, (3) weather conditions at the time of exposure may vary, and (4) the possible inability to determine reliably an individual's threshold to motion sickness in a single controlled session.

Mechanical Devices

Swings have been used most frequently to determine motion sickness susceptibility. The general procedure is to swing an individual for 20 minutes to half an hour (less, if he vomits), and to grade the response of the man according to the number of symptoms he displays and perhaps the time at which the symptoms appear. Shortcomings of the swing and other devices have rested upon the inability to describe adequately the relation of the stimulus time-intensity to incidence.

Studies of the validity of swings to evaluate susceptibility to motion sickness have been made using as criteria: (1) subsequent reaction to sea- or airsickness, (2) comparison of the responses of men disqualified for duty because of motion sickness with those of men demonstrated to be resistant, and (3) correlation of questionnaire histories of motion sickness with responses to the swing.

In his summary of the literature, McIntyre states, "The balance of evidence at present available indicates that proper use of the swing test as a screen should enable a substantial reduction to be made in the incidence of primary airsickness . . ." (**79**). He pointed out, however, that failure to control the test properly, such as failure to fix head position during the swing test, may produce spurious results. Hemingway (**47**) points out that it is necessary to repeat swing tests on all individuals who are sick on the first occasion. In this manner, the men showing no adaptation to motion sickness might be separated from the men who are somewhat susceptible but who would adapt if given sufficient exposures.

According to Table III, taken from Hemingway's data (**47**), if all men who reacted to the swing had been eliminated, about 27 percent, or 116, of the total group would have been screened. Of these, 62 men had a history of motion sickness and 54 did not. The efficiency of this screening can be estimated from the fact that of the 33 percent of the men with positive histories of motion sickness, one half could have been eliminated on the swing test but at the expense of one "good" man for every susceptible man.

A relation between history of motion sickness and symptoms induced by a "Roll Rocker" machine was reported by Morton *et al.* (**85**). It would appear from the results in Table IV that the roll rocker had a somewhat greater predictive efficiency than the swing. About two-thirds of the men with a history of motion sickness displayed symptoms on the *machine* whereas a little less than half the men with a positive history had symptoms of motion sickness on the *swing*. An even closer relation was reported by Alexander *et al.*, between symptoms induced by a vertical accelerator and a questionnaire history of motion sickness (**7**). In this study, the vertical accelerator "sickness rates" for subjects grouped according to their previous history of motion sickness were as follows: "susceptibles, 45 percent; intermediates, 24 percent; non-susceptibles, 14 percent." The authors point out that their findings were based upon controlled laboratory conditions and might, therefore, be higher than would be obtained under operating conditions in the field.

TABLE III

Correlation of Swing Reactions with History of Motion Sickness

Swing Sickness	History of Previous Motion Sickness					
	No		Yes		Total	
	N	%	*N*	%	*N*	%
Immune	231	54.2	79	18.5	310	72.7
Mild	34	8.0	25	5.9	59	13.9
Severe	20	4.7	37	8.7	57	13.4
	285	66.9	141	33.1	426	100.0

TABLE IV

Correlation of Reactions to the "Roll Rocker" and History of Motion Sickness

Severe Symptoms With 30 Min. on Machine	History of Motion Sickness		
	Resistant	Unknown	Susceptible
Yes	1	34	28
No	25	74	13
Total	26	108	41

TABLE V

Correlation of Reaction to a Swing Test and to a Short Cruise at Sea

Swing Reaction	Sea Reaction			
	Unaffected	Nauseated	Vomited	Total
Unaffected	21	5	8	34
Doubtful nausea	5	1	6	12
Definite or marked reaction	7	3	10	20
Total	33	9	24	66

Seasickness and swing sickness were correlated in a study (**54**) of 66 men who were tested on a swing and also taken to sea (Table V). The investigators concluded that the swing was of little practical value as a preselection test for seasickness. It would seem that they were unduly conservative in their interpretation, since of the 32 men who had some positive reaction on the swing test, 20 also had a positive reaction to seasickness.

Of a total of 1,000 swing tests in the AAF, 11 percent of the non-airsick flyers vomited, in contrast to 65 percent of the airsick eliminees (**49**). While Hemingway points out that this is not perfect correlation, it appears to be high enough to be of value in screening procedures.

Some of the objections to the determination of motion sickness susceptibility on a swing or other device rest upon the difficulty of retesting the men to get a reliable measure. In one of the few quantitative studies of the reliability of the swing test, Jasper *et al.* (**58**) point out that 19 percent become sick on only one of two test sessions. The correlation obtained between tests 1 and 2 was 0.84 for 77 cases. It is likely that more precision could be achieved in the determination of a threshold by improving the criteria of onset of symptoms.

Questionnaires

Questionnaire forms have been developed which require the person to indicate the various forms of motion that have ever made him "sick" and if so, how frequently and how severely (**7, 15**). A questionnaire history of motion sickness has high inherent validity in that a large sample of exposure to motion is surveyed. It may be more useful to know whether a man has ever been sick on all the forms of transportation than to know his reaction to a single test on a device. Questionnaires and devices are not mutually exclusive, however, and the use of both can be combined in an effective screening procedure.

Reliability. Questionnaires have been criticized more on the basis of reliability than upon their intrinsic validity and ability to predict susceptibility to motion sickness. Four factors may detract from the reliability of a motion sickness questionnaire: (1) failure by the individual to record motion sickness, e.g., because of zealousness to get into service, (2) overemphasis of motion sickness, e.g., by a malingerer, (3) poor memory of personal experience with motion, and (4) misinterpretation of symptoms, e.g., a man may never associate a headache, dizziness or other minor symptoms on a ship or plane with motion sickness.

While these factors may be important in individual cases, studies (**18**) of questionnaire-gathered histories of motion sickness indicate that their reliability is high ($r = 0.90$). No significant difference was found for the average response on a motion sickness questionnaire when given to Navy recruits along with other procedures at a boot camp, (a) when they had to sign the questionnaire, and (b) when they were told it was for research purposes only, could not affect them in any way, and no names were required.

Judgments of susceptibility to seasickness made by individuals themselves were found to correlate quite highly with judgments by their shipmates, although self judgments of susceptibility tended to indicate slightly higher susceptibility than judgments by others (**18**).

Validity. The relation between a history of motion sickness and susceptibility to motion sickness on devices, e.g., swings, vertical accelerator and a complex roll rocker has been shown to be significant (**7, 49, 85**). Susceptibility to seasickness, as demonstrated by 150 men aboard a destroyer escort, was found to be related to their history of previous motion sickness (**20**). This relation was sufficiently high to reduce seasickness if the men had been screened on the basis of the questionnaire. About half of the crew who were highly susceptible to seasickness were among the highest 10 percent of the questionnaire scores.

Interviews

It has been suggested that psychiatric screening would reduce the number of men likely to be incapacitated by some form of motion sickness (**22, 64, 101**). Similarly it has been thought that the individual who is persistently motion sick is a potential psychiatric casualty. Although considerable has been written about emotional factors in motion sickness, there is little quantitative evidence about the reliability and validity of psychiatric or non-specific psychological interviews in detecting men likely to be very motion sick. More information is needed before even tentative conclusions may be drawn about the usefulness of psychiatric interviews in screening for motion sickness.

Exposure to Motion of Actual Environment

In many instances, the process of training may afford useful information about the susceptibility of the individual. Not all personnel are exposed to motion during their training, however, since certain technical skills are required before they are assigned to duty aboard ship. It is not uncommon to have personnel trained at shore stations for many months before assigning them to the fleet. While it may be desirable to have a short period of sea duty before extensive training, this procedure is difficult on a large scale. The possible usefulness of a preliminary evaluation of susceptibility to motion sickness by exposure to the actual working environment, has yet to be explored.

Labyrinthine Tests and Tests of Balance

Responses to rotation on the conventional Bárány Chair test appear to have little relation to susceptibility to motion sickness. Results on 644 aviation cadets indicated that qualification or non-qualification on the Bárány Chair test (according to the medical criteria of the test) was unrelated to symptoms of motion sickness on a swing test (**76**). Similarily, post-rotational nystagmus was reported to be no different for a group of men sent to shore duty because of seasickness than for a random group of men (**17**).

Symptoms produced by injection of "hot" or "cold" water into the external ear, i.e., caloric vestibular tests, are not closely related to susceptibility to airsickness (**38, 79, 85**). The bulk of evidence suggests, therefore, that there is little, if any, relation between the kinds of measurements furnished by caloric or rotation tests and susceptibility to motion sickness. There is similarly no evidence that susceptibility to motion sickness is a reflection of a hypersensitive vestibular apparatus, at least as indicated by measurements on the caloric and Bárány Chair tests.

Studies were carried out on a group of men highly susceptible to seasickness to discover if their balance was different from normals (**19**). Two tests were used: (1) the rail-walking test, which required the individual to walk a thin wooden rail while holding his hands behind his back, and (2) the ataxiagraph test, which recorded the amount of body sway while standing erect, feet together, under eyes open and eyes closed conditions. The susceptible men were no different from normals in ability to walk a thin rail. On the ataxiagraph, however, significantly greater body sway was reported for the susceptible group. This finding may be a reflection of some defect in the tonic vestibular mechanisms which exercise a reflex control on posture.

Emetic Drugs

Proneness to vomiting has been tested by the use of drugs which reproduce many of the symptoms of motion sickness (**16, 35**). Results of the administration of such drugs, e.g., neostigmine and apomorphine, to men highly susceptible to seasickness were compared with results on a random group of men (**16**). One of the drugs, neostigmine, did not differentiate the groups, although a significantly greater number of symptoms was displayed by the susceptible men in response to apomorphine. The difference was not so great as to suggest that susceptibility to motion sickness is simply a reflection of the "threshold" of the vomiting center or its inhibition. In view of the small number of cases, the findings must be regarded as only suggestive. Furthermore, in the particular group studied, the somewhat greater disposition to display symptoms in response to the drug may not have existed as an original predisposing factor and may have resulted from experience with seasickness (**16**).

Acclimatization

Attempts have been made to raise the "threshold" to motion sickness by slightly increasing the duration of a swing test each day for a week or more. In this manner, it is possible to render persons more resistant to the motion of a swing so that they can tolerate longer periods of swinging without the appearance of symptoms (**70, 79**). The significant issue, however, is the amount and duration of transfer that occurs between such adaptation to a swing and the motion of a ship or an airplane. The literature reviewed by McIntyre gives little evidence of transfer of adaptation on a swing to the motion of airplanes or ships (**79**). Furthermore, the adaptation to the swings themselves is transitory, lasting perhaps a week. Also relevant is a report of an attempt at adaptation to a swing among a group of men grounded because of persistent airsickness (**25**). There was little evidence of adaptation in this group. Two men that did show some adaptation during daily swinging were again motion sick on the swing after several weeks of rest. These findings corroborate the anecdotal evidence that naval personnel susceptible to seasickness must readapt on each cruise, i.e., regain their "sea legs" on each cruise.

Drugs

Administration of motion-sickness-preventive drugs to active personnel is a more complex problem than similar medication for passengers. Undesirable side effects of drugs such as possible interference with vision, sleepiness, or dryness of the throat, are of greater importance in men expected to function at peak psychomotor efficiency than in passengers who might be given heavy sedation. Despite rigorous criteria with respect to medication of active personnel, several remedies have been found to be relatively free from side effects that would preclude their use. No consistent physiological rationale for drug control can as yet be advanced. The literature on the use of drugs in motion sickness is voluminous, larger perhaps than all other aspects of motion sickness combined. The accomplishments have not been commensurate with the attention devoted to drugs, and many of the early reports are merely anecdotal accounts with neither experimental controls nor critical analysis of results.

Tyler's report (**111**) of the effects of nine different drug combinations contains most of the relevant information on the effectiveness of drugs in preventing motion sickness. In this work, the drug dosages were determined by the amount of a drug that could be given without interfering with the efficiency of the men. Another important feature of these studies was the use of both a non-medicated control group and a lactose (placebo) control group. Sixty experiments were conducted on over 15,000 unselected young men engaged in amphibious training operations off the California coast. Hyoscine alone or in combination with other drugs was found to be effective in preventing motion sickness. An average of 60 percent protection was afforded by hyoscine or hyoscine combinations when the control rates were as high as 52 percent and half of these had severe symptoms. Barbiturates were less effective than hyoscine when the general incidence of motion sickness was high. In these studies, the medications were given one to three hours before embarking. In situations where the medication could not be given until evidence of motion sickness was already present, it is likely that the benefit would be less because of the interference with absorption of the drugs.

Hyoscine appears to be the favored drug in control of motion sickness, and when given orally (0.6 mg.) about one hour before exposure to motion, can significantly reduce the incidence of motion sickness (**51, 103, 111**).[1]

Indoctrination

Acquainting all personnel with the known facts about motion sickness may help to minimize the problem, particularly by dispelling some of the attitude of expectancy surrounding the first cruise aboard ship. While it is difficult to prevent the experienced hands from exploiting the ignorance of the greenhorn, the possession of the facts of motion sickness by the neophyte may avoid useless anxiety and being made the butt of pranks designed to aid the appearance of motion sickness rather than prevent it. It is well known that airsickness may be easily produced on first flights by maneuvers; the simple expedient of requiring the pilot rather than the student to clean up

[1] Since this chapter was written, a new remedy (Dramamine) has been reported (*Science*, 1949, **109**, 359–360). The initial studies appear to place the drug within the range of effectiveness of the better motion-sickness remedies (**111**), and further studies of its relative effectiveness are indicated, as well as investigations of undesirable side-effects of the drug.

the cabin after flight can result in a remarkably lower incidence of first flight sickness.

Dissemination of facts about motion sickness in pamphlet form may improve the general attitude of service personnel toward motion sickness so that it may become regarded as an essentially normal process. The desired result would be to have the susceptible person looked upon merely as differing in degree in the same manner as visual or auditory acuity or even stature, and not as "weak" or "queer."

Engineering Design

From the point of view of controlling motion sickness alone, few principles of design can as yet be stated, although, in general, the larger the vehicle the lower the periodic accelerations. Another complicating factor is high speed, which decreases the interval between accelerations and decelerations and also increases the peak amplitude. Speed is also important by reason of associated sudden starts and stops and by its relation to a rough traveling medium.

Increase in size, arrangement of seating at the center of gravity, or rotation and provision for viewing of the horizon are factors that might be utilized in reducing motion sickness. Placement of seating and arrangement of windows in vehicles may have an important effect on motion sickness, since informal interviews with persons highly susceptible to motion sickness indicate that motion sickness occurs more frequently in the rear seats of automobiles than in the front and is greatest for the rear seats of limousine-type vehicles. In many instances, it is impossible to permit seating at the center of rotation and also vision of the horizon, e.g., aboard ships and airplanes. The usefulness of an artifical horizon for such conditions should be studied.

Control of Auxiliary Factors

Significant reductions in the incidence of motion sickness can be made by proper control of some of the auxiliary factors attending the exposure to periodic motion. The most important of these factors appear to be vision and head position. The latter is, in a sense, a primary factor since it determines the direction of the accelerations imparted to the vestibular apparatus; in the present context, however, it is convenient to discuss it as an auxiliary factor.

Vision

The critical factor in vision appears to be the orientation supplied by viewing the stable environment, e.g., the horizon or landscape. There are three general conditions of visual orientation which have been studied concomittantly with exposure to motion: (1) eyes open viewing the stable environment, (2) eyes closed, and (3) eyes open, vision restricted to the moving environment, e.g., eyes open below deck or eyes open in a covered swing. The incidence of motion sickness appears to be lowest under condition (1) and greatest under condition (3). Blindfolding subjects increases motion sickness on a swing. When a swing is enclosed with a "cabin" and the subjects keep their eyes open, a further increase is noted (**71, 79**). These results do not agree with those of Spiegel (**106**) who reported that viewing a light in the moving environment reduced motion sickness induced by rotational accelerations. The difference in type of acceleration may account for the apparent disagreement. Further work should be directed at analyzing this issue.

The importance of inhibition of motion sickness through visual orientation is revealed by the data of Tyler (**111**). His studies compared the incidence of motion sickness in men standing in landing craft viewing the horizon, with men crouching below the gunwales so that they could not see the horizon or other ships; head position was comparable.[2] In the visually "ori-

[2] Although head position was comparable, free movement of the head was more difficult in the crouching position. Should acceleration in one plane be important, this difference in free head movement may have some relation to the difference in incidence of motion sickness noted by Tyler.

ented" groups there was an average sickness rate of 11.7 percent and a range of rates of 5–19 percent with 2.2 percent severely sick. The "non-oriented" group had an average rate of 30 percent motion sickness, and a range of 25–42 percent with 10.1 percent severely sick. Under the conditions of the experiments, the effectiveness of visual orientation was as great as that afforded by drug medication. This finding may be regarded as reliable, since it was demonstrated on observations of over 2,100 men in 24 experiments.

From these results it might be concluded that the use of a visual artificial horizon below deck might reduce the incidence of motion sickness. This approach was actually adopted by Tyler in an unpublished study in which he mounted an artificial horizon near the operating stations of below deck personnel (**113**). The results of the study suggested that the artificial horizon did have the predicted effect of reducing motion sickness; but the investigator regarded the findings as tentative. This type of experiment definitely needs further development because of its possible practical value as well as its theoretical importance.

Magladery (**68**) has indicated that the early British gliders had small port-hole-type windows that markedly restricted the vision of the horizon by glider personnel. It was thought that this restriction of vision potentiated the appearance of motion sickness and it was recommended that subsequent glider models be provided with larger windows; this change was believed to have appreciably reduced the incidence of motion sickness.

Although studies show an effect of vision on motion sickness, it is yet not clearly settled whether vision can exercise both a facilitating and an inhibiting effect under different conditions. It is worthy to note that there have been no successful attempts to produce symptoms of motion sickness by vision alone (**62**). A moving picture taken from the subjects' position on a swing failed to produce any symptoms of motion sickness in subjects shown the film for over 30 minutes (**85**).

Reduction in motion sickness achieved by lying down is primarily the result of changing the position of the head, i.e., the vestibular apparatus, with respect to the applied acceleration. In a series of experiments on head and body position, it was found that regardless of body position, a relation existed between head position and the direction of the changes in acceleration (**55, 56, 78, 107**). The incidence of motion sickness in these studies was greatest if the changes in acceleration were perpendicular to the plane through the external auditory meatus and the lateral canthus of the eye; lower rates of motion sickness were noted when the changes in acceleration were parallel to this plane. Such findings give objective support to confirm the impression that sufferers from motion sickness secure relief from lying in bed, beyond the mere effect of resting alone.

These results also afford evidence that the effect of acceleration upon the viscera and blood column is not an important factor, if present at all, in the production of motion sickness. Indirect confirmation is also contained in the observation that no relation existed between the amount of visceral movement in individuals and susceptibility to motion sickness (**69**).

Numerous factors have been mentioned in the folklore of motion sickness that are now known to be largely irrelevant. The use of special diets, conditioning exercises, pre-embarkation cathartics, and other clinical devices have been minimized in the face of experimental evidence (**114**). Fatigue, odors, illness, and other factors have also been mentioned in connection with motion sickness. Without specific information available, it is difficult to estimate their importance, although in view of the trend toward defining motion sickness with greater specificity, it is reasonable to view the importance of such possible determinants with suspicion. It is worth noting here that any importance attached to cerebral anoxia resulting from interference with blood flow in motion sickness must be minimized (**32**).

Air temperature has long been thought to be important in motion sickness, because of the subjective effect of feeling better when going into the open air topside. Aboard ship this is obviously complicated by changes in vision now known to be significant in inhibiting motion sickness. Constant conditions for vision exist in the standard swing test; and an analysis of results of 1000 swing tests during various temperatures throughout the year ranging from 0°C to 40°C, indicate little effect of temperature on the incidence of motion sickness (**42**). Slightly greater motion sickness was found for temperatures of 36–40°C, but in this temperature range only a few tests were run, so that the results are equivocal.

The time before or after meals, or the time of day one is exposed to motion apparently has little effect upon the occurence of motion sickness (**6**). Relevant to this finding of the lack of importance of meals is the additional negative finding that "nothing in the roentgenologic appearance of the stomach before swing testing was found to correlate with whether the subject became 'swing-sick' or not (**75**)".

Psychogenic Factors in Motion Sickness

Probably the most confused issue is the role of such factors as suggestion, fear, and personality traits in the production of motion sickness. Although such factors were emphasized in early reports, there is now a trend away from regarding susceptibility to motion sickness as a manifestation of neurosis or similar defect. Historically, it is apparent why a "functional" or psychogenic point of view was popular. For one thing, the stimulus (acceleration) and the sense organ affected (the vestibular apparatus) were vaguely understood. Also, the person being made motion sick is not aware of the effect of acceleration, since the vestibular apparatus does not have the degree of cortical representation of other organs of special sense. Furthermore, lacking a single-factor explanation, the early clinicians who worked on the problem were prone to accept a loose and non-specific functional interpretation which made susceptibility to motion sickness, in a sense, equivalent to a form of neurosis.

From the results of research on motion sickness in animals, it is clear that motion sickness can occur without unusual psychological predisposing factors. The mere coexistence of neurosis with susceptibility to motion sickness need not point to a relation between them since the high incidence of neurotic complaints in contemporary society insures the frequent association of a neurotic personality and susceptibility to motion sickness in the same individuals.

Fear and Motion Sickness

The principal emotional factor that has been emphasized in motion sickness is fear. In extreme fear or anxiety, symptoms resembling those of motion sickness may be seen, and it has been suggested that fear is an important element, if not *the* critical factor, in airsickness (**10**). It should be realized, however, that the symptoms of motion sickness resemble those produced by a wide variety of conditions that have little in common psychologically. Fear has been most stressed in airsickness, although even in this instance, the hypothesis must face the embarrassing fact that experienced pilots may become airsick if traveling as passengers (**119**). In a study of naval personnel, it was found that there was no relation between fears of drowning, explosion, and fire aboard ship, and susceptibility to motion sickness (**18**).

The only experimental investigation of the role of fear in motion sickness employed an electric shock given to persons as they were rotated in a chair. Individuals receiving anticipated electric shocks during rotation had a greater incidence of motion sickness than control subjects who did not receive shocks during rotation (**121**). In this study, an individual was regarded as sick if he requested that the motion be stopped before 12 minutes. This criterion of sickness is complicated by the fact that an electric shock is an annoyance by itself, and individuals may have requested that the motion be stopped not only because they felt ill, but because they wished to avoid further shocks. Of interest is the finding that there was no relation between "neurotic" tendency and susceptibility to motion sickness on the rotating chair. Within the limitations of the study it would appear that, if fear is an important factor, it operates independently of neurotic personality qualities. Of 14 individuals in the study who had no previous history of motion sickness, only 4 were made sick upon rotation *with* shock. Of 14 other subjects who had a history of motion sickness, 6 were made sick on the chair *without* the electric shock. It would thus appear that the basic susceptibility of an individual was relatively more important than his response to the fear-inducing stimulus. The finding that electric shock potentiates the appearance of motion sickness would be most relevant to first flights or the first day aboard ship. In sea- or airsickness, the repeated exposure to the motions of the ship or plane would reduce the element of fear through familiarity (**43**).

Neurotic or Personality Maladjustment

One of the early studies on motion sickness reported that neurotic children who tend to be car sick have an emotional association with motion that has a significant effect upon the appearance of symptoms (**28**). This report prompted others to investigate the possibility of a relation between personality and susceptibility to motion sickness. Many of the studies of this type are subject to criticisms that grow out of the difficulties of obtaining adequate control subjects and also for failure to recognize alternative explanations. For example, men examined in a Naval Hospital because of severe seasickness shortly after they had been discharged from a ship are likely to present an unusual picture because of the loss of weight, which in some cases amounted to 30 pounds or more during the past war. The impact of the physical debility plus the realization by the man that he has not been able to perform his job adequately can easily prejudice the examiner into assuming he is dealing with a psychological and constitutional misfit.

Observations on men sent to a Naval Hospital because of seasickness led one investigator to conclude that a large proportion of the men severely affected by motion sickness have neurotic complications (**100, 101**). Another study of men sent to shore because of severe seasickness was based upon observations obtained after the men were reassigned to shore duty. In the latter case, the men had an opportunity to gain back lost weight and the chance to perform a shore job (**17**). Although there was a higher percentage of individuals with psychological defects than might be expected by chance, the proportion was much lower than that expected if severe seasickness were solely an expression of emotional or neurotic defect. It was also found that a history of motion sickness correlated more closely with susceptibility to seasickness than did neurotic tendencies as measured by a psychosomatic complaint questionnaire prepared for use in psychiatric screening. This result was based upon the study of several groups of naval personnel that totalled over 1,500 individuals, including men aboard ship, men who had completed tours of sea duty, and new recruits (**18**). Fraser (**33**) reported similar results in a study of 80 consecutive cases of airsickness washouts. In this sample there were 36 normal, 32 "neurotic," and 12 doubtful cases of personal adjustment. Also worthy of note was the additional observation that, of 20 men grounded for psychoneurotic disorders, none manifested airsickness, despite the fact that two men had such severe anxiety states that they had been reported to lie on the floor of the aircraft incapable of activity. Such results prompt the suggestion that MacPhee and Pennington analyze their findings in greater detail since they merely reported that "emotional instability" was present in 100 men with a history of motion sickness (**64**).

A report of the Flying Personnel Research Council of Great Britain on 120 men eliminated from duty because of airsickness indicated that only 9 cases of neurosis were found and 51 cases of neurotic predisposition (**108**). In only 19 of the 120 cases of airsickness disability was neurotic predisposition thought to play a role. In a group of 30 consecutive navigator cadets eliminated for severe airsickness, the "emotional" factor was believed to be predominent in 15 cases (**22**). The evidence gives support to the general conclusions of Symonds and Williams which may well serve as a summary to this section, with the reader substituting the term "motion sickness" for "airsickness": "When a man is suspended for airsickness at any stage of training, the cause is usually motion sickness uncomplicated by psychological factors. Psychological factors, either neurosis, neurotic predisposition or faulty morale—may contribute by lowering the physiological threshold for tolerance of motion, or by reducing the man's ability or willingness to endure symptoms, but psychological factors are seldom of major importance" (**108**).

Suggestion

Suggestion has been thought to play a role in motion sickness in both a positive and negative manner. If one person becomes motion sick aboard ship or in an airplane, presumably others may become sick by witnessing the distress of the ill person. Conversely, however, if one is given a pill with no active ingredient, a placebo, and told it will cure motion sickness, then by suggestion alone amelioration or prevention of symptoms has been thought to occur. Tyler (**111**) studied the incidence of motion sickness in three groups of personnel: (1) no medication, (2) placebo or dummy pill, and (3) active drugs. The incidence of motion sickness was as great in the placebo group as in the untreated group, whereas the active drugs significantly reduced the incidence. If suggestion were important, the opposite results would have been obtained, i.e., the untreated group would have had a significantly higher incidence than the placebo group and the active drugs would not have had a significant advantage over the placebo treated group. Tyler's results are especially conclusive, since they are based on several experiments and were conducted by independent observers under a variety of sea conditions.

Physiological Factors in Motion Sickness

Three areas of physiological data may be distinguished: (1) conditions of the environment which alter the physiological state of the body and the appearance of motion sickness, (2) the sequence of physiological events involved in motion sickness, and (3) the internal physiological states that are relatively independent of the environment but may affect the susceptibility of the individual to motion sickness. The latter two areas have been studied by observations on persons before, during, and after motion sickness and by detailed studies of persons of different degrees of susceptibility in an effort to analyze the physiological basis for individual differences. Such studies have attempted to discover why, in the face of the same pattern of acceleration, one person becomes motion sick while another displays no signs of being ill. An important distinction must be made between those factors which are necessary for the production of motion sickness and those which are not only necessary, but which also determine the relative susceptibility of an individual. Thus, we do not know the contribution of the threshold of the vestibular apparatus to individual differences, although we do know that these sense organs are necessary for the appearance of motion sickness. It is conceivable that the sense organs could be but faithful transmittors of the changes in environmental acceleration and that the appearance of motion sickness depends upon the relative facilitation or inhibition of stimuli leading to the excitation of various centers in the brain.

In motion sickness there is a decrease in the rhythmic contractions of the stomach and intestine, i.e., decreased gastric tone and peristalsis (**75**). However, there are no characteristics of the gastrointestinal tract which are related to susceptibility to motion sickness. In a detailed study of persons with various degrees of motion sickness susceptibility, there were found no anatomical or functional aspects of the gastrointestinal tract which were associated with susceptibility to motion sickness (**75**). It would appear that responses of the gastrointestinal tract in motion sickness are not significantly influenced by individual differences in the characteristics of these organs. It is likely that conflicting results in the past have been due to the high residual incidence of gastrointestinal defects in men chronically seasick for a protracted period (**100**). Such findings could result from continuous seasickness rather than existing as a predisposing condition, since negative results were obtained from examinations of 10 such men after several months of shore duty (**19**).

Studies of the brain as related to motion sickness have been of two types, (a) electroencephalograph records of humans, and (b) extirpation experiments on animals. Abnormal brain waves were reported to occur with slightly greater frequency in very susceptible persons (**79**). Several subsequent studies, however, have indicated little or no relation between cortical brain activity in the electroencephalograph and susceptibility to seasickness (**19, 60, 85**). In one of these studies, the response to hyperventilation was also checked and no relation to motion sickness was found (**60**).

The integrity of portions of the cerebellum appears necessary for the appearance of motion sickness although much, if not all, of the cortex of the cerebrum may be removed without eliminating motion sickness in dogs (**12, 13**). Susceptible dogs subjected to surgical removals of the cerebellum indicated that a limited area (nodulus and uvula) of this structure is critical for the production of motion sickness. This result was not merely due to interference with the vomiting mechanism, since the dogs could be stimulated to vomiting with an emetic drug, apomorphine.

Many of the detailed physiological observations have failed to indicate significant differences between susceptible and non-susceptible persons (**8, 48, 85**). Negative findings for such factors as cardiovascular responses, blood chemistry, blood gases, respiratory rate, and general medical evaluation are important as they point to motion sickness as a more specific response than formerly realized.

Indications for Future Research

Although present knowledge of motion sickness can be used to reduce the magnitude of the problem, there are many areas needing research. When the basic question is asked, "Why do people become motion sick?", the inadequancy of present knowledge becomes apparent. It is a little surprising that few theories of the mechanisms involved have been advanced in recent years. Even teleological speculation about motion sickness is very unsatisfying, for it is difficult to perceive the reasonableness of motion sickness as a response. Control of motion sickness will be handicapped until we can state with certainty, for example, the action of acceleration upon particular sensory receptors, the nature of the afferent nerve impulses, the neural pathways in the central nervous system, and have a more complete knowledge of the facilitating and inhibiting influences upon the processes.

Continuation of research on the action potentials of the sensory nerves of the vestibular apparatus should be fruitful. Such studies as the recording of sensory impulses during various patterns of acceleration should indicate whether a qualitatively different kind of afferent discharge arises under conditions of motion sickness. This work can also indicate whether the labyrinthine receptors are mere passive recorders of the stimulating conditions or whether the structural and functional aspects of these receptors are such that they play an important role in differences in susceptibility. Direct electrical stimulation of the vestibular portion of the eight cranial nerve is also a means of determining the importance of temporal pattern of stimulation. Further tracing of the vestibular pathways in the brain is necessary to strengthen knowledge about the role of various portions of the brain in motion sickness.

Effort should be made to derive a quantitative expression for the threshold stimulus and to establish the stability and reliability of such measurements. In contrast to the single threshold value, another value representing an adaptability coefficient is needed. Once a reliable measure of adaptability is established, studies can be carried out on the optimal conditions for creating adaptation. Perhaps such experiments can be made utilizing the analogies of conditioning experiments, since adaptation resembles experimental extinction rather than sensory adaptation, according to present indications. Since adaptation or extinction of motion sickness at present seems rather specific, attention might be given to experimental methods of generalizing the phenomenon.

Visual factors in motion sickness need further analysis, since artifacts of head position and movements have not as yet been satisfactorily ruled out. The use of the artificial horizon in this connection appears particularly fruitful. Such instrumental control of visual orientation may be used to supply "true" orientation or "conflicting" orientation to determine whether or not vision may act both in the role of an inhibiting as well as a facilitating influence.

A few well-controlled experiments are needed to examine more closely the relation between susceptibility and personality factors. An attack can be made by exposing to controlled motion individuals known to be severely upset psychologically, e.g., institutionalized psychoneurotics. The use of independent judgments in assaying personality factors is particularly important and for that reason groups of persons already categorized as to personality traits should be employed as subjects. This would minimize the effect of the experimental bias so often pointed out as being important in this area. Additional work on the role of fear and suggestion should be done to establish the range of effect that might result from these variables on motion sickness. Fear of the situation involving motion could be induced by a variety of techniques, including ingoctrination lectures and post-hypnotic sugdestion.

Applied research might be set up in situations where follow-up records are possible. Such studies would permit the application of several tests of susceptibility before exposure to motion, e.g., aboard ship. Analysis of such information would permit a retrospective estimate of the efficiency of the criteria. Follow-up studies would give a more accurate account of the incidence and adaptability to motion sickness than now available.

Search for drugs to reduce motion sickness should be continued, although studies should also be made of drugs which facilitate the appearance of motion sickness. It may prove as valuable from theoretical grounds to know drugs which facilitate motion sickness as to know the inhibitors.

Summary

Motion sickness consists of a pattern of symptoms including sweating, nausea, and vomiting that occurs as a result of exposure to periodic acceleration. More than 50 percent of the general population may be made motion sick. Such sickness is observed on airplanes, swings, trains, and in numerous

other situations in which transient or periodic accelerations are encountered. Animals, other than man, share the predisposition to motion sickness, and dogs in particular have been found to be useful for experimental studies of the problem.

Motion sickness is primarily the result of stimulation of the vestibular apparatus of the inner ear by acceleration. Destruction of the vestibular apparatus or its afferent nerve fibers results in immunity to motion sickness. Recent research defines motion sickness with greater specificity, and non-labyrinthine factors are now considered to be less important.

Continued motion sickness, e.g., on a long cruise at sea, may require hospitilization of individuals due to protracted vomiting. Although the number of cases who are unable to show any adaptation to motion is small, these individuals represent a serious problem in the military services.

It is improbable that a totally resistant group could ever be successfully screened from the population at large because of the large proportion of individuals who may be made motion sick. Susceptibility to motion sickness may be tentatively regarded as a general factor for all situations involving acceleration. This quality of susceptibility may be utilized to screen personnel in military or other situations where a more resistant population is desired. With available means of detecting susceptibility to motion sickness, e.g., questionnaires and swings, a significant reduction could be made in the number of unadaptable or the very susceptible persons being assigned to critical tasks in military service.

Of all auxiliary factors attending the appearance of motion sickness, visual orientation and position of the head with respect to the direction of the applied acceleration appear to be most important. The magnitude of reduction of motion sickness by control of head position or vision is, in general, larger than the protection afforded by drug therapy.

Drugs have been found to be effective in the reduction of motion sickness, and protection up to 60 percent has been reported. Suggestion accompanying the use of medications is not regarded as significant, since studies have shown that the incidence of motion sickness is about the same in untreated as in placebo groups, whereas groups receiving active medication have a significantly lower incidence.

Although there are marked individual differences in susceptibility to motion sickness, the basis for such differences are not apparent. Few measurable traits either physiological or psychological have been found to be significantly related to susceptibility. Individual differences in personality traits do not contribute to an important degree to variations in susceptibility. Motion sickness is decidedly not a psychogenic affliction, although there are undoubtedly individuals in whom unusual emotional states may facilitate the appearance of symptoms, once the primary conditions of acceleration are met.

Although present knowledge of motion sickness can be applied to reduce the problem of motion sickness, several aspects need further research. Continued investigation into the basic nature of motion sickness with attention to refining the criteria of sickness and improving the reliability of the various measurements should result in still further gains in control of the problem.

References

1. Anonymous. Critical study of seasickness remedies. *Royal Nav. med. Bull.*, 1943, No. 4, 3–6.
2. Anonymous. Airsickness in military aviation. Summary Report No. 2, War Department, Army Air Forces, No date.
3. Anonymous. A psychological and physiological survey of aircrew enlistments in the Royal New Zealand Air Force. NRC, New Zealand File, Fitness for Aircrew, November 1944.
4. Agin, L. J. Airsickness and related psychosomatic complaints. *Air Surgeon's Bull.*, August 1945, **2**, 235–236.

5. Alexander, S. J., Cotzin, M., Hill, C. J., Ricciuti, E. A., & Wendt, G. R. Studies of motion sickness; effects of sickness upon performance. *J. Psychol.*, 1945, **20**, 31–39.
6. Alexander, S. J., Cotzin, M., Hill, C. J., Ricciuti, E. A., & Wendt, G. R. Studies of motion sickness; incidence of sickness at various hours of day. *J. Psychol.*, 1945, **20**, 19–24.
7. Alexander, S. J., Cotzin, M., Hill, C. J., Ricciuti, E. A., & Wendt, G. R. Studies of motion sickness; prediction of sickness on vertical accelerator by means of motion sickness history questionnaire. *J. Psychol.*, 1945, **20**, 25–30.
8. Alexander, S. J., Helmick, J. S., Taylor, J. H., & Wendt, G. R. Studies in motion sickness, Series C (Experimental comparison of various autonomic responses in individuals susceptible and nonsusceptible to motion sickness). CAA Div. of Res., Report No. 66, 19–29, 1946.
9. Alexander, S. J., Helmick, J. S., Hill, C. J., & Wendt, G. R. Studies in motion sickness, Series C (A second experimental comparison of autonomic responses in individuals susceptible and nonsusceptible to motion sickness). CAA Div. of Res., Report No. 66, 31–46, 1946.
10. Armstrong, H. G. *Principles and practice of aviation medicine.* Baltimore: Williams & Wilkins, 1943. Pp. 246–256.
11. Army Medical Library. *Bibliography of motion sickness.* (microfilm) Washington, D. C., Photoduplication Service, Army Medical Library, 1942.
12. Bard, P. Physiological investigation of cause and nature of motion sickness. Final Report, OEMcmr-257, CAM Report No. 485, September 30, 1945.
13. Bard, P., Woolsey, C. N., Snider, R. S., Mountcastle, V. B., & Bromiley, R. B. Delimination of central nervous mechanisms involved in motion sickness. *Fed. Proc. Amer. Soc. exp. Biol.*, March 1947, **6**, 72.
14. Benedict, E. B., & Schwab, R. S. Gastroscopic studies in naval personnel with chronic seasickness. *New England J. Med.*, 1945, **233**, 237.
15. Birren, J. E., Fisher, M. B., & Stormont, R. T. Evaluation of a motion sickness questionnaire in predicting susceptibility to seasickness. Research Project X-278, Report No. 2, Naval Medical Research Institute, August 15, 1944.
16. Birren, J. E., Stormont, R. T., & Pfeiffer, C. C. Reactions to neostigmine and apomorphine as indication of susceptibility to seasickness. Research Project X-278, Report No. 4, Naval Medical Research Institute, February 26, 1945.
17. Birren, J. E., & Morales, M. F. Observations on men highly susceptible to seasickness with remarks on periodic motion of ships. Research Project X-278, Report No. 5, Naval Medical Research Institute, July 17, 1945.
18. Birren, J. E., & Fisher, M. B. Further studies on the prediction of susceptibility to seasickness by a motion sickness questionnaire. Research Project X-278, Report No. 6, Naval Medical Research Institute, October 8, 1945.
19. Birren, J. E. Psychophysiological studies of motion sickness. Ph.D. thesis, 1947. On file, Northwestern Univ. Library.
20. Birren, J. E., & Fisher, M. B. Susceptibility to seasickness; a questionnaire approach. *J. appl. Psychol.*, 1947, **31**, 288–297.
21. Bond, D. D. A psychiatric analysis of forty subjects made sick by a swing. A.A.F. School of Aviation Medicine, Project 149, Report 1, Randolph Field, Texas, July 31, 1943.
22. Bond, D. D. A study of the emotional factors involved in a group of thirty consecutive navigation cadets in the process of elimination for severe airsickness. A.A.F. School of Aviation Medicine, Project 136, Report 1, Randolph Field, Texas, April 17, 1943.
23. Brooks, M. The etiology of seasickness. *Med. Rec.*, N. Y., 1939, **150**, 23–26.
24. Brown, B. R., Brett, J. R., & Howlett, J. G. Swing sickness and a history of motion sickness. Second Report, Ass. Comm. Aviation Med. Research, Appendix F., Report No. 1, July 22, 1943.
25. Brown, G. L., McArdle, B., & Magladery, J. W. Interim report on clinical investigations into airsickness. Report FPRC, No. 410–A.
26. Camis, M. *The physiology of the vestibular apparatus.* Oxford: Clarendon Press, 1930.
27. Claremont, C. A. The psychology of seasickness. *Psyche*, Lond., 1930, **11**, 86–90.
28. Conn, J. H. Psychiatric study of car sickness in children. *Amer. J. Orthopsychiat.*, 1938, **8**, 130–141.
29. Dorcus, R. M. Influence of physiologically effective doses of epinephrine on vestibularly induced nausea. Reports, Div. Anthrop. & Psychol., NRC, November 1942.

30. Duff, I. F. Medical study of the experiences of submarines as recorded in 1,471 submarine patrol reports in World War II. Research Division, Bu. Med. & S., Navy Department, Washington, D. C., No Date.
31. Foucar, F. H., Gordon, B. S., & Kaye, S. Fatal poisoning from motion sickness preventive. *J. Amer. med. Ass.*, July 20, 1946, **131**, 971–972.
32. Fraser, A. M. Preliminary study of the effect of mild anoxia upon the incidence of swing sickness. Proc. Exec. Ass. Comm. Aviation Med. Res., Appendix D, March 22, 1944.
33. Fraser, W. H. Emotional factor in airsickness. Reports, FPRC, RAAF, FR 87, April 1944.
34. Gibson, W. C., Manning, G. W., & Kirkpatrick, B. Effect of active vestibular training on motion sickness. (Summary only), Proc. Conf. Motion Sickness, NRCC, Appendix 1, June 16, 1943.
35. Goehring, H. D., & Schwab, R. S. The use of the swing, the injection of prostigmine, and caloric stimulation of the labyrinth in the estimation of susceptibility to chronic seasickness. U. S. Naval Hospital, Chelsea, Mass., No Date.
36. Green, D. M. Airsickness in bomber crews. *J. Aviat. Med.*, 1943, **14**, 366–372.
37. Griffith, C. R. The organic effects of repeated bodily rotation. *J. exp. Psychol.*, 1920, **3**, 15–46.
38. Hallpike, C. S., & Fitzgerald, G. M. Observations on the pathology of airsickness. Report FPRC, No. 334, July 1941.
39. Helmick, J. S., & Wendt, G. R. Studies in motion sickness, Series C (Circulatory and respiratory responses to cold and to breath holding in individuals susceptible and nonsusceptible to motion sickness). CAA, Div. of Res., Report No. 66, 1–17, 1946.
40. Hemingway, A. Cardiovascular changes in motion sickness. *J. Aviat. Med.*, 1945, **16**, 417–421.
41. Hemingway, A. Cold sweating in motion sickness. *Amer. J. Physiol.*, 1944, **141**, 172–175.
42. Hemingway, A. Environmental temperature and swing sickness. *J. Aviat. Med.*, 1946, **17**, 86–88.
43. Hemingway, A. Incidence of airsickness in cadets during their first ten flights. AAF School of Aviation Medicine, Project 170, Report 5, Randolph Field, Texas, January 29, 1945.
44. Hemingway, A. Motion sickness among air crew personnel. AAF School of Aviation Medicine, Project 126, Report 1, Randolph Field, Texas, March 3, 1943.
45. Hemingway, A. Motion sickness history and results of the swing test of one hundred and seven airsick eliminees from flight training. AAF School of Aviation Medicine, Project 170, Report 2, Randolph Field, Texas, September 7, 1943.
46. Hemingway, A. Relationship of airsickness to other types of motion sickness. *J. Aviat. Med.*, 1946, **17**, 80–85.
47. Hemingway, A. Results on 500 swing tests for investigating motion sickness. AAF School of Aviation Medicine, Project 31, Report 2, Randolph Field, Texas, November 25, 1942.
48. Hemingway, A. Survey of research on the problem of airsickness in the Army Air Forces, AAF School of Aviation Medicine, Project 381, Report 1, Randolph Field, Texas, April 20, 1945.
49. Hemingway, A. Selection of men for aeronautical training based on susceptibility to motion sickness. *J. Aviat. Med.*, 1946, **17**, 153–163.
50. Hill, A. B. Report on correlation between swing-sickness and airsickness. Report, Comm. Airborn Troops, MRC-BPC 42/57 ABT 8, June 5, 1942.
51. Hill, I. G. W., & Guest, A. I. Prevention of seasickness in assault craft; report of experiments under tropical conditions. *Brit. med. J.*, 1945, **2**, 6–11.
52. Hoff, E. C. *A bibliography of aviation medicine.* Springfield: Thomas, 1942.
53. Hoff, P. M., Hoff, E. C., & Fulton, J. F. *A bibliography of aviation medicine—a supplement.* New Haven: Yale Univ., 1942.
54. Holling, H. E., McArdle, B., & Trotter, W. R. Observations on the correlation of swing sickness and seasickness. Report, Comm. Airborne Troops, BPC 42/110.
55. Howlett, J. G., Wardill, T. E. M., & Brett, J. R. Effect of position on the incidence of swing sickness. (Summary only). Proc. Ass. Comm. Aviation Med. Research, Appendix F, Report 4, July 22, 1943.
56. Howlett, J. G., Wardill, T. E. M., & Brett, J. R. Effect of position on the incidence of swing sickness. (Supplementary Report). Proc. Ass. Comm. Aviation Med. Research. Appendix F, Report 5, July 22, 1943.
57. Howlett, J. G., & Brett, J. R. Speculation on the mechanism of utricular response to stimulation in motion sickness. Proc. Ass. Comm. Aviation Med. Research, Appendix F, Report No. 6, July 22, 1943.

58. Jasper, H. H., Battista, A., Bornstein, M., & Noble, R. H. Reliability of the swing test of motion sickness. Proc. Ass. Comm. Naval Med. Research, Appendix N-2, October 7,1943.
59. Joekes, A. M. Correlation of swing sickness with airsickness. FPRC, Report No. 475.
60. Lindsley, D. B., & Wendt, G. R. Studies in motion sickness, Series A (An investigation into the relationship of the electroencephalogram to motion sickness susceptibility). CAA, Div. of Res., Report No. 40, 25–34, 1944.
61. Lindsay, J. R., Oppenheimer, M. J., Wycis, H. T., & Spiegel, E. A. Receptor apparatus of the vestibulovasomotor reaction. *Arch. Otolaryng.*, 1945, **42,** 257–266.
62. Livingston, R. C. Apparatus for investigating visual reactions in relation to airsickness. FPRC, Report No. 463.
63. Maaske, C. A. Production of motion sickness by swinging in the prone position. War Dept., Air Forces, ENG-49-696-44, June 30, 1943.
64. MacPhee, H. M., & Pennington, L. A. Identification of cases of chronic motion sickness during military screening examinations. *J. soc. Psychol.*, 1946, **23,** 61–64.
65. Magladery, J. W. Interim report of investigations into airsickness in gliders. IV. Effect of repeated swinging on susceptibility to airsickness. Report, Comm. Airborne Troops, BPC 42/56 (c).
66. Magladery, J. W. Interim report on investigations into airsickness in gliders. V. Effect of special physical training on susceptibility to motion sickness. Report, Comm. Airborne Troops, BPB 42/56 (d).
67. Magladery, J. W. Investigations into airsickness in gliders. Report, Comm. Airborne Troops, MRC-BPC 42/56.
68. Magladery, J. W. Personal communication.
69. Mainland, D. Influence of body position on the level of abdominal viscera in relation to seasickness. Min. Comm. Naval Med. Research, Appendix C, NRCC, January 28, 1942.
70. Manning, G. W. Acclimatization to swing sickness. Proc. Ass. Comm. Aviation Med. Res. Appendix F, February 25, 1944.
71. Manning, G. W., & Stewart, W. G. Effect of position on the incidence of swing sickness. Proc. Exec., Ass. Comm. Aviation Med. Res., Appendix F, April 18, 1943.
72. McDonough, F. E. Airsickness during an airborne infantry maneuver. CAM Report No. 102, November 16, 1942.
73. McDonough, F. E. Airsickness in the airborne infantry. AAF School of Aviation Medicine, Project 93, Report 2, Randolph Field, Texas, November 27, 1942.
74. McDonough, F. E. Airsickness in navigation training. AAF School of Aviation Medicine, Project 165, Report 1, Randolph Field, Texas, July 5, 1943.
75. McDonough, F. E., & Schneider, M. The effect of motion on the roentgenographic appearance of the stomach and small bowel. *Gastroenterol.*, 1944, **2,** 32–45.
76. McDonough, F. E., & Thorner, M. W. Regulation Bárány and swing tests in navigation cadets. AAF School of Aviation Medicine, Project 100, Report 1, Randolph Field, Texas, December 23, 1942.
77. McEachern, D., Morton, G., & Lehman, P. Seasickness and other forms of motion sickness. *War Med.*, 1942, **2,** 410–428.
78. McIntyre, A. K. Effect of head position on susceptibility to motion sickness. Report, Comm. Flying Personnel Research, FPRC, RAAF, FR-86, November 1943.
79. McIntyre, A. K. Motion sickness, present status of research. Report, Comm. Flying Personnel Research, RAAF, FR-91, June 1944.
80. McIntyre, G. Correlation between swing sickness and airsickness. Report, Comm. Flying Personnel Research, RAAF, FR-67, No Date.
81. McNally, W. J. The otoliths and the part they play in man. *Laryngoscope,* 1944, **54,** 304–323.
82. McNally, W. J., & Stuwart, E. A. Physiology of the labyrinth reviewed in relation to seasickness and other forms of motion sickness. *War Med.*, 1942, **2,** 683–771.
83. Michael, W. H. Suggested attack on seasickness. *Proc. U. S. Naval Inst.*, 1947, **68,** 1747.
84. Morales, M. F. Asynchrony of labyrinthine receptors as a physical factor in motion sickness. *Bull. Math. Biophys.*, 1946, **8,** 147–157.
85. Morton, G., Cipriani, A., & McEachern, D. Mechanism of motion sickness. *Arch. Neurol. Psychiat.*, 1947, **57,** 58–70.
86. Noble, R. L. Differences in susceptibility to motion sickness in dogs and the effect of treatment with barbiturates. Proc. Ass. Comm. Army Med. Research, **8,** 26–27, Appendix D, Vol. I, November 1943.

87. Noble, R. L. Methods of assaying motion sickness preventives on dogs. *Canad. J. Res.*, Sect. E., 1945, **23,** 226–234.
88. Noble, R. L. Motion sickness. Proc. Ass. Comm. Army Med. Research, NRCC—10–11, November 1944.
89. Noble, R. L. Observations on various types of motion causing vomiting in animals. *Canad. J. Res.*, Sect. E. 1945, **23,** 212–225.
90. Noble, R. L. Treatment of experimental motion sickness in humans. *Canad. J. Res.*, Sect. E, 1946, **24,** 10–22.
91. Noble, R. L., Sellers, E. A., & Best, C. H. The treatment of motion sickness. *Canad. Med. Ass.*, 1947, **56,** 417–424.
92. Office of Scientific Research and Development. Bibliography of airsickness and seasickness. Report No. 82, November 1942.
93. Park, J. Airsickness in gliders. FPRC, Report 510.
94. Park, J. Correlation between swing sickness and airsickness and history of motion sickness. FPRC, Report 485, August 24, 1942.
95. Plaut, A. B. J. Psycho-somatic approach to disabilities in South African Air Force flying personnel. Analysis of Central Medical Establishment Disability Boards, S. A. Report No. 28, March 20, 1945.
96. Pumphrey, R. E., Kuh, J. R., Johnson, M. A., & Day, A. L. Unusual conditions aggravated by seasickness. Report of 2 cases. *U. S. naval med. Bull.*, Wash., 1946, **46,** 742–743.
97. Rubin, H. J. Airsickness in a primary Air Force training detachment. *J. Aviat. Med.*, 1942, **13,** 272–276.
98. Rushmer, R. J., & Bond, D. D. The hyperventilation syndrome in flying personnel. *War Med.*, 1944, **5,** 302.
99. Rushmer, R. F. Overventilation in subjects during the swing test. AAF School of Aviation Medicine, Project 133, Report 1, Randolph Field, Texas, May 1, 1943.
100. Schwab, R. S. Chronic seasickness. *Ann. int. Med.*, 1943, **19,** 28–35.
101. Schwab, R. S. Chronic seasickness; neurological, psychiatric and naval aspects. *Trans. Amer. neurol. Ass.*, 1942, **68,** 25–28.
102. Sjöberg, A. A. Experimentelle Studien über den Auslösungsmechanismus der Seekrankheit. *Acta oto-laryng.*, Suppl. 14, 1931, 1–136.
103. Smith, P. K. Effectiveness of some motion sickness remedies in preventing airsickness in air force navigation students. *J. Aviat. Med.*, 1946, **17,** 343–345.
104. Sokalchuk, A. Blood pressure in seasickness. *Fed. Proc. Amer. Soc. exp. Biol.*, 1947, **6,** 207.
105. Spiegel, E. A. Effect of labyrinthine reflexes on the vegetative nervous system. *Arch. Otolaryng.*, 1946, **44,** 61–72.
106. Spiegel, E. A., Oppenheimer, M. J., Henny, G. C., & Wycis, H. Experimental production of motion sickness. *War Med.*, 1944, **6,** 283–290.
107. Stewart, W. G., & Manning, G. W. Effect of tilting the head back on the incidence of sickness on the 4 pole swing (summary only), Proc. Ass., Comm. Aviation Med., Research, **1,** Appendix F, Report No. 7, July 22, 1943.
108. Symonds, C. P., & Williams, D. Investigation of psychological disorders in flying personnel. The psychological aspects of airsickness. FPRC, Report No. 593, September 1944.
109. Trotter, W. R. Seasickness. *Practitioner*, 1946, **157,** 45–49.
110. Tuttle, A. D., & Wendt, G. R. Studies in motion sickness, Series B (Airsickness among one hundred eighty-nine airline stewardesses and its relationship to previous history of motion sickness). CAA, Div. of Res., Report 60, 54–73, 1946.
111. Tyler, D. B. Influence of placebo, body position, and medication on motion sickness. *Amer. J. Physiol.*, 1946, **146,** 458–466.
112. Tyler, D. B. Motion sickness and physiological problems peculiar to parachutists and other airborne troops. Final report, OEMcmr-282, May 31, 1946.
113. Tyler, D. Personal communication.
114. Tyler, D., & Bard, P. Motion sickness. *Physiol. Rev.* (In press)
115. Wendt, G. R. Studies in motion sickness, Series A (A note on an unsuccessful effort to investigate the effects of temperature on vestibularly induced nausea). CAA, Div. Research, Report 40, 35–45, 1944.
116. Wendt, G. R. Studies in motion sickness, Series B (Airsickness among seventy-one student pilots and fifteen instructors and its relationship to previous history of motion sickness). CAA, Div. Res., Report 60, 75–80, 1946.
117. Wendt, G. R. Studies of motion sickness. XI. The test-retest reliability of a motion

sickness history questionnaire. Final report, OEMcmr-524.

118. WENDT, G. R. Studies on motion sickness. XII. Frequency of airsickness in pilots during elementary and primary naval flight training. Final report. OEMcmr-524.

119. WENDT, G. R. Of what importance are psychological factors in motion sickness? *J. Aviat. Med.*, 1947, **19**, 24–32.

120. YACORZYNSKI, G. K. Differential modification of post-rotational nystagmus of normal pigeons and pigeons with cerebral lesions, with comparisons of cerebro-cerebellar relationships. Ph.D. thesis, 1937, On file, Northwestern Univ. Library.

121. ZWERLING, I. Psychological factors in susceptibility to motion sickness. *J. Psychol.*, 1947, **23**, 219–239.

CHAPTER 19

MOTION SICKNESS: PHYSICAL CONSIDERATIONS REGARDING ITS ETIOLOGY

MANUEL F. MORALES

University of Chicago

INTRODUCTION

It is evident that certain features of the motion sickness problem are easier to describe physically than are others. For instance, even with present-day methods we can expect to obtain a fairly good knowledge of the time course of the forces exerted by various craft on their passengers. We shall show below that the afferent neural response which these forces evoke from the receptors is also understandable in physical terms. When we come to study the ensuing behavior of the central nervous system, however, we are rapidly forced back to purely qualitative descriptions, and indeed this part of the chapter will add nothing to what has already been said regarding central phenomena. This disparity in levels of information makes presentation difficult, and calls for special caution on the part of the reader. Specifically, he must not jump to the conclusion that the biophysical approach is acceptable in studying receptor phenomena but futile in studying central phenomena; still worse would it be to infer from the larger space allotted to the description of receptors that motion sickness is a purely peripheral problem. Finally, he should exercise tolerance in demanding immediate practical results, for the progress reviewed below is much more by way of understanding than of solving the motion sickness problem.

THE MOTION OF NAVAL VESSELS

The motion of a vessel can be regarded as a superposition of oscillations with respect to certain coordinates, e.g., the vertical distance from the center of mass to a horizontal reference plane, the angle which a mast makes with the vertical, etc. These oscillations are executed in response to periodic forces applied on the vessel by the waves. If one knew the mathematical form of the applied forces and the dynamic constants of the vessel (its geometry, distribution of masses, etc.) it should be possible to deduce the observable oscillations. As with many such problems, however, solution by conventional methods is hopeless (**31, 22**), and one takes refuge in approximations which are frequently of the crudest sort. For the purposes of estimating the forces which a vessel exerts on its passengers it has been the custom to use the following semi-empirical method. By comparatively simple means (e.g., a bubble gauge inclinometer and a stopwatch) one may measure the average periods and amplitudes of the oscillations. *Assuming* that these oscillations are simple harmonic, one may then readily interpolate to obtain the entire time course of the oscillations, and also calculate velocities and accelerations at any instant. Thus, if t stands for time, and if the observed average amplitude and period of an oscillation of the coordinate x are x_0 and T respectively, one assumes that

$$\ddot{x} = -(2\pi/T)^2x, \quad (1)$$

where the superscript dot refers here and elsewhere to differentiation with respect to time. From this it follows that the coordinate is given at any time by

$$x = x_0 \sin (2\pi/T)t \quad (2)$$

(taking $x = 0$ when $t = 0$). The velocity and accelerations are simply obtained by differentiating equation (2); in the resulting expressions, putting the trigonometric factor equal to unity gives the *maximum* velocity and acceleration. If x is an angle (in radians), then at a distance R from the axis of rotation the linear (tangential) displacement is obtained by multiplying equations (1) and (2) through by R. This method of computation was used in connection with motion sickness by Sjoberg (**34**), who considered the motion of a vessel to be the super-position of three oscillations (Fig. 1): rolling, or rotation in the transverse plane; pitching, or rotation in the longitudinal plane; and scending, or vertical translation of the center of mass normal to the horizontal plane. The axes of the rotations and of the translation intersect approximately at the center of the vessel near the water line.

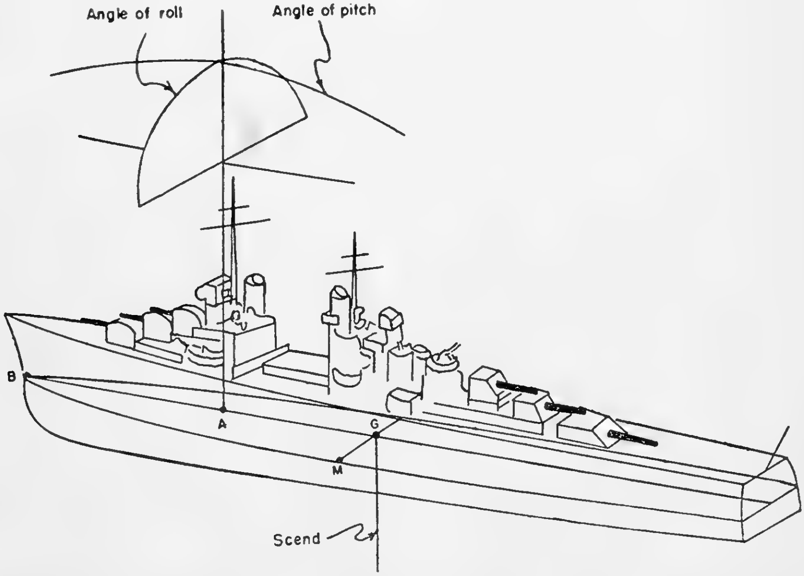

Fig. 1. Schema to illustrate the three important oscillatory motions of a vessel, and the positions at which the accelerations in Table I have been computed.

Sjoberg (**34**) studied the contributions of rolling and pitching in a "passenger vessel of average size," paying special attention to the linear accelerations thus generated (for reasons to be discussed below). He concluded that the maximum angular accelerations which could be expected would be about 5° per sec., and the maximum linear accelerations about 0.5 g. Morales (**23**) has made similar computations for several classes of naval vessels, employing sallying data.[1] In this study the maximum linear accelerations (obtained on the assumption that all components add) were calculated at three positions on the vessel (Table I and Fig. 1); an average of these values for each vessel gives a rough idea of the upper limit of forces to which personnel may be subjected, and also gives some notion of the relative capacity of various classes of vessels for producing these forces. It may be noted parenthetically that the "average" maximum acceleration for the various classes parallels at least roughly the sickness-inducing characteristics of the classes. It is also well to mark for future reference that most of the accelerations lie

[1] Kindly furnished by the Design Stability Section of the Bureau of ships, Navy Department.

well below .8 g. For certain craft—landing craft and destroyers—the maximum angular accelerations are considerably higher than Sjoberg's limit, being in one extreme caes (LCI, roll) 49° per sec.^{2}

As already mentioned, the foregoing calculations, although based on experimental measurements of periods and amplitudes, depend ultimately upon the assumption of simple harmonic motion, and it is therefore important to obtain continuous records of displacement in order to check the assumption. No doubt the difficulty of constructing sensitive instruments which will resolve the motion along two planes accounts in part for the scarcity of data. Incidental to other objectives, Morales and Birren (**24**) attempted such measurements aboard a battleship, using the level and cross-level angles registered by an AA gun computer trained on 000. Although their study was purely preliminary, it indicated that the technique was satisfactory, and that extensive routine observations could thus be made on all classes of vessels carrying stable elements (destroyers and up, and some submarines; the newer submarines, unfortunately, do not carry these instruments). It appeared from these records that, although the angular motion was not precisely simple harmonic, it was, on the other hand, of such a type that in *moderate* seas its coarse features could be described fairly well by the simple harmonic assumption. This was not true of records taken during a storm; indeed a storm could be described just as well by the erratic character of the record as by the shift to higher average amplitudes. The observed periods of pitch (moderate seas)—4–5 seconds—were about the same as those in sallying, but the periods of roll—8 seconds—were shorter; consequently the average maximum linear accelerations were greater than those computed for the battle ship in Table I. However, they are still well within the .8 g maximum.

TABLE I

Some Characteristics of the Motion of U. S. Naval Surface Craft

Vessel Type	Period of Roll (Seconds)	Amplitude of Roll (Degrees)	Period of Pitch (Seconds)	Amplitude of Pitch (Degrees)	Average Acceleration
Destroyer	9.5	25	5	5	.26 g
LCI	4.5	25	3	5	.21 g
LST	5.0	15	4	3	.19 g
Destroyer Escort	8.0	25	5	5	.14 g
Heavy Cruiser	12.0	20	7	4	.14 g
Carrier	16.0	15	7	4	.13 g
Light Cruiser	12.0	20	7	4	.13 g
Escort Carrier	13.0	20	7	4	.12 g
Battleship	14.0	12	6	3	.12 g
Transport	14.0	25	7	4	.08 g

This table is constructed entirely from unrestricted information. The periods and amplitudes are average values kindly furnished the author by the Bureau of Ships, Navy Department. The "average acceleration" is the arithmetic mean of the roll and pitch contributions at three points aboard the vessel: at the bow on the water line, abeam of midships on the water line, and high on the foremast.

A more detailed and complete study using similar apparatus was undertaken at about the same time by Deacon at the Admiralty Research Laboratory (**8**). Deacon undertook measurements aboard a small naval auxiliary (about 150 ft. long, weighing 800 tons), along a coast where simultaneous shore-based measurements of the waves were possible. Although his investigations were directed at a different problem, viz., studying wave motion using ship motion as an index, they probably contain the best data available bearing on the present problem.

Deacon frequency-analyzed his records and thus obtained for any one experimental condition the *spectrum* of pitch, roll, and wave periods. This is the most useful representation for data of this type and no doubt should be adopted in all future work. When coursing parallel and anti-parallel to the wave direction, the spectrum of pitch-periods was not coincident with the spectrum of wave periods (short periods being more prominent in the former), but an overall Doppler effect was quite evident at various speeds of the ship. On the other hand the modal periods of roll were hardly affected by course and speed.[2] The spectrum of pitch periods was very broad (a more or less flat distribution between 5 and 17 seconds).

From the two studies just mentioned, especially the last, it appears that extended records of ship motion can be taken rather easily, and that, as a rough approximation, simple harmonic motion is acceptable. Deacon's findings with respect to roll (and, to some extent, pitch) suggest that whereas the wave conditions are a modifying factor, the motion of a vessel is governed chiefly by its own dynamic constants.

All measurements mentioned thus far have referred to surface vessels; data on submarines are even more scarce. It has been stated that roll periods of submarines are considerably greater than those of surface craft; also, that surface swell causes rolling at depth as great as 100 ft. (an extreme case has been reported in which a boat rolled through 15° at a depth of 150 ft.). The author has not, however, found any systematic, reliable data regarding roll at various depths. Recently, in response to a questionnaire issued to 15 fleet-type submarines, the Bureau of Ships has collected some data on motion near the surface under various wave conditions (Table II).[3] It would appear from the data, contrary to common belief, that neither periods nor amplitudes are of unusual character. There was in the data a suggestion that longer wave lengths led to longer periods, but the precision of these observations does not warrant any conclusions other than the tentative one that submarines near the surface do not roll or pitch "abnormally" and hence estimates of the sort deducible from Table I can be supposed to apply.

TABLE II

Some Characteristics of the Motion of U. S. Submarines

	Pitch	Roll
Average amplitude (deg.)	1–3 (Mode, 2)	1–17 (Mode, 8)
Maximum amplitude (deg.)	1–4 (Mode, 3)	1–27 (Mode, 17)
Period (sec.)	3–7 (Mode, 5)	5–11 (Mode, 8)

Summary

The following conclusions may be submitted by way of summarizing features of ship motion which bear on the motion sickness problem. It should be emphasized, however, that the paucity of accurate measurements renders such conclusions very tentative.

1. Three to seven seconds probably includes all the important pitch periods encountered in naval vessels; 5 to 15 seconds probably includes all the roll periods. This appears to be true of submarines as well as of surface craft.

2. The simple harmonic assumption cannot be considered a very good approximation

[2] The range was only slightly shifted.

[3] These data were made available to the author through the courtesy of Mr. Charles Moore, Code 444, Bureau of Ships, Navy Department.

to the actual motion, but probably gives correct orders of magnitude for such derived quantities as average linear accelerations, etc.

3. The maximum linear accelerations encountered aboard naval vessels are about .8 *g*; the average (over various points and various classes) accelerations are probably .3 *g* or less. Maximum angular accelerations can be as high as 49° per sec.[2]

4. Continuous records of roll and pitch can be obtained using gyroscopic equipment of the fire-control type.

5. For any one vessel there are unique spectra of roll and pitch characteristics, and it is essential that these spectra be determined from the continuous records.

6. The action of waves modifies to some extent (e.g., Doppler effect) the periods and amplitudes of the motion, but the characteristics of the ship seem to play the dominant role in governing the motion.

Mechanics and Neural Response of the Receptor Systems

Prior to a discussion of the response of the labyrinthine receptors it is useful to have in mind a simple model of their mechanical arrangement (Fig. 2). One may imagine each sacculus and utriculus to consist of a saucer-shaped base lined with sensory epithelium (macula). Overlying the epithelium is attached a membrane of greater specific gravity (otolith) (**40**). It appears (**1**) that, in these receptors, effective stimulations are those which tend to pull the otolith away from the macula; one may imagine that in such cases the hairs of the sensory epithelium are stretched. There is some evidence (**20**), however, indicating that these organs also respond continuously under static conditions. The second class of receptor organs is that of the semicircular canals (Fig. 2). Any one canal may be regarded as a doughnut-shaped rigid container filled with fluid (endolymph).[4] Extending into the canal, at the enlarged region (ampulla), there is a projection (the cupula) hinged at its base (crista), in which base lies the sensory epithelium. Any accelerations—either acting directly on the cupula or transmitted by the endolymph—which cause a deflection of the cupula are adequate stimulation (**39**); presumably these deflections stretch the hairs of the sensory epithelium imbedded at the cupular hinge.

[4] That the fluid outside of the central canal plays any role seems to be ruled out by the experiments of Dohlman (**12**).

It has already been stated that in the sacculus and utriculus as well as in the

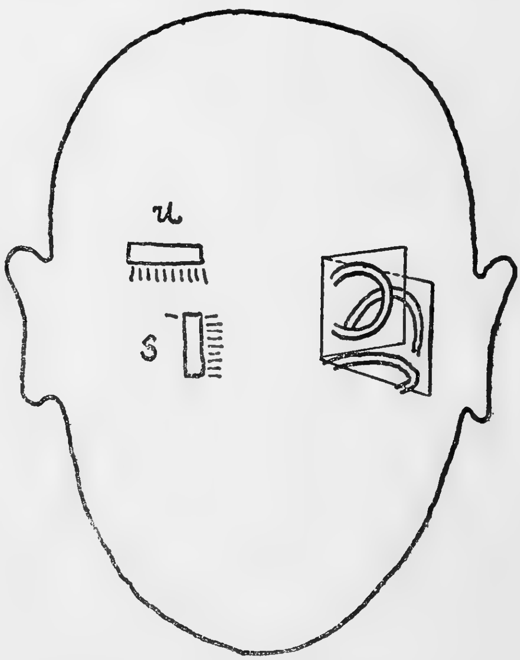

Fig. 2. Schema of the utriculus (u) and sacculus (s) of the right side, and the canals of the left side.

canals, stimulation results from the deflection of an elastic (and also viscous) structure. One must, however, know the nature of the relation between the extent of deflection and the neural response of the receptors. In the case of the canals, this relation has been elucidated by a number of classical studies. Steinhausen (**39, 11**) first clearly demonstrated on the pike that cupular deflection was quantitatively associated with signals emitted from the receptors—using as a criterion of activity the nystagmic response.

Three subsequent studies—all in substantial accord—have considered in greater detail the relation between known motions and the direct frequency response in the nerve fibers coming from the cristae (of a horizontal canal): Ross (**30**) on the frog, Lowenstein and Sand (**16**) on the ray, and Adrian (**1**) on the cat. It seems legitimate to assume that their general observations will also apply to man. Taking Lowenstein and Sand's results as representative, one may conclude that: (a) There is, under static conditions, a steady frequency of response from the crista. (b) On deflection to one side the frequency increases, and on deflection to the other side the frequency decreases.[5] (c) Over a matter of 20 to 30 seconds, at least, "adaptation"—i.e., a decay with time, of the frequency response when the deflection is held constant—is negligible.[6] (d) To a fairly good degree of approximation[7] the percent change in the frequency is proportional to the angle of deflection (**23**). These are all facts of which use will be made presently.

The nerve response of the sacculus and utriculus has not been studied so widely. Adrian (**1**), using a considerably more indirect (but necessarily so, for dissection is much more difficult in the cat) procedure, has reported response curves of the sacculus in which the frequency of discharge appears to rise slightly faster than linearly with angle of static inclinations of the head (for angles from 0°–40°). This curvature is in the opposite direction from that which would result from a simple geometric factor (sine of the angle of inclination), and therefore probably represents a true non-linearity in the response. As in the case of the canals, however, the response shows little adaptation. From experiments in which the animal's head was accelerated up and down, Adrian concluded that acceleration downward probably stimulated the utriculi (whose otoliths under these conditions would be pulled away from their maculae). Accelerations upward probably stimulated the dorsal lobes (which face downward) of the sacculi for the same reasons. Accelerations sideways probably stimulate the sacculi, and do so asymmetrically (**12, 15**), that is, only on one side of the head is the otolith pulled away from the maculae.

In the case of the sacculus and utriculus there is no indication, as there is in the canals, that a single macula responds to acceleration in both directions normal to itself, but instead that it reacts only when the otolith is pulled away from it (see, however, the interpretation of swing experiments discussed below).

To test quantitatively a hypothesis about the mode of action of the labyrinthine receptors requires first an analytic formulation of the hypothesis. Such a formulation, once justified experimentally, is useful in diverse ways to be indicated presently.

Information of the proper sort to test hypotheses about the sacculus and utriculus is so meager that formulations of the action of these receptors must remain for the time being purely speculative. We shall here merely note that if the otolithic system deflects like a damped vibrator, and that if one restricts the problem to deflections small enough so that the afferent nerve fiber frequency, ν, is proportional to the deflection, then ν satisfies the equation,

$$\ddot{\nu} + 2\lambda\dot{\nu} + \kappa^2\nu = f(t) \qquad (3)$$

where 2λ and κ^2 are numbers proportional

[5] This was an unequivocal result in the Lowenstein and Sand experiments; the existence of a two-directional sensitivity was problematical in Adrian's work, and absent in Steinhausen's. In Mowrer's work on the pigeon (to be cited below) there was again evidence that a receptor on one side was sensitive to rotation in both directions.

[6] Jonason and his co-workers (**14**) have reported an interesting effect on guinea pigs of much more prolonged stimulation, viz., degenerative changes in the labyrinth as a result of exposure to angular accelerations of 39.4° per sec.^{2} for 100–200 hrs.

[7] The approximation here involved is, in all probability, that the angle of deflection must be small (say, less than 25°).

respectively to the viscosity and elasticity of the system, and $f(t)$ is proportional to the instantaneous normal component of the force acting to deflect the otolith from its equilibrium position relative to the macula. This force may arise in various ways, e.g., one may reach into the labyrinth and pull away an otolith from its macula, either directly or (in the case of the sacculus) by the centrifugal force generated in a rotation; or, one may accelerate the macula, by accelerating the entire cranium, in which case the inertial force of the otolith causes the same drawing away action. In all these cases, once $f(t)$ is specified, equation (3) is readily integrable to give $\nu(t)$ as a function of the parameters 2λ and κ^2. From a suitable analysis of experiments in which $f(t)$ was known and $\nu(t)$ was observed, 2λ and κ^2 may be calculated, and from them $\nu(t)$ for any $f(t)$. So far, however, such experiments have not been done, and the justification of equation (3) is largely on structural grounds.

In the case of the canals, elementary mechanical considerations lead to an equation like (3), the assumption being that the cupula deflects as a damped vibrator, and that the change in afferent fiber frequency is proportional to the deflection. The forcing function $f(t)$ now represents the sum of the torques brought about by the motion imposed on the cranium. As an illustration we may consider the stimulation of a vertical canal by rotation about a distant axis in its plane—a situation not unlike that in pitching and rolling. The motion imposed on the walls of the canal can be resolved into (two) translations and a rotation, and the effects may be considered separately. That a pure *translational* acceleration can twist the cupula out of its resting position is theoretically certain, for not only will it generally give rise to an inertial torque, but also it will generate an antiparallel pressure gradient in the endolymph (and therefore a torque, unless the angular and endolymphatic densities are identical).[8] This is an important objection to the assumption of many writers (see **21**) that the canals are not involved in detecting linear accelerations. A pure rotation gives rise to two inertial torques on the cupula, and—by virtue of the friction between canal walls and endolymph—to an endolymph thrust equivalent to a torque. The experiments of Lowenstein and Sand (**16**), in which $\nu(t)$ was observed and the known $f(t)$ were the torques of pure rotations, are described quite well by equation (3). This circumstance therefore permits a calculation of 2λ and κ^2, and from them a prediction of $\nu(t)$ in the course of oscillating rotation, i.e., ship-like motion (**23**).

Crude as the analytic studies to date have been, they have nonetheless provided certain important results: they have ruled out persistent endolymph flow as the explanation of persistent neural responses such as nystagmus (**33**), and have shown (**38, 39**) that such phenomena are fully explained by the high $2\lambda/\kappa^2$ ratio for cupular deflection. The numerical evaluation of 2λ and κ^2 has further suggested (**23**) that the frequency of the oscillations imposed by ships may be a frequency which renders the cupular deflections just 90° out of phase with the imposed displacement, and therefore presumably with instantaneous receptors like the eye. Whether this maximal asynchrony is an etiologic factor in motion sickness is still for future experimentation to decide. Perhaps the most useful contribution which the analytic model of labyrinthine receptors makes to the study of motion sickness, however, is the suggestion of a quantitative measure of the stimulating effectiveness of an imposed disturbance. We shall describe this measure briefly now,

[8] Sjoberg (**34**) has studied this effect both theoretically and in models. More recently, Morgan, Summers, and Reimann (**20**), have given an interesting derivation of the expressions for the gradient in the case of half-canals (actually, the so-called "semi-circular" canals are hydraulically complete canals, i.e., the liquid column forms a ring).

and return to its application in the next section.

We have said that, granting the model of receptor action described above, integration of equation (3) for a given $f(t)$ yields $\nu(t)$, or, in the event that there is a rest frequency, ν_0, it yields $\nu - \nu_0$; in any case one has a measure of the *instantaneous* rate of stimulation of the central nervous system by the receptor. The *average* rate of stimulation over a time interval, $t_2 - t_1$, we shall take to be σ, where,

$$\sigma^2 = 1/(t_2 - t_1) \left\{ \int_{t_1}^{t_2} (\nu - \nu_0)^2 \, dt \right\}. \quad (4)$$

For *periodic* imposed motions of period, T, we have the special case of equation (4),

$$\sigma^2 = (1/T) \left\{ \oint (\nu - \nu_0)^2 \, dt \right\}, \quad (5)$$

where $\oint$ indicates integration over one complete cycle. In the next section we shall evaluate σ^2 by integrating equation (3) and substituting the resulting value of ν into equation (5).

To this point we have dealt with the mechanical problems posed by the action of the receptor mechanism when the imposed motion is especially simply directed with respect to the receptor in question. Very little attention has been paid in the past to the geometrical arrangement of the receptors in the head. This is excusable if the imposed motion is a pure vertical acceleration and the receptor supposed to be involved in the process is the utricle. In the case of arbitrarily directed linear accelerations, and especially in the case of rotations such as those attending roll and pitch, there is no question but that the imposed accelerations must be resolved into appropriate effective components. Resolution into linear horizontal and vertical components is simple, but it is interesting to note that the components of the acceleration which act upon the linear receptors in the "adapting head"—that is, a head which is undergoing no rotation even though the ship is rolling or pitching—have a rather complicated time dependence, certainly a very different one from that of the angular displacement (**23**). A more complicated case is that of the resolution of an imposed angular velocity into components normal to the planes of the six semicircular canals (**41**). Specifically, one has a vector angular velocity $\vec{\omega}$, whose instantaneous components in some fixed rectangular reference, (X, Y, Z) frame are known. It is desired to know the components of $\vec{\omega}$ in another rectangular (ξ, η, ζ) frame which moves relative to the first in pure rotation. In the case in point the ξ, η, ζ axes are taken normal to the planes of the canals. The most convenient coordinates to describe the position of this moving frame are the Euler angles (see Osgood's *Mechanics*, **29**, p. 216), ψ, θ, φ. When the (human) head is still, we may take the X–Y plane as the horizontal, the Y axis as the intermeatal line, and the positive Z axis along the upward vertical. Referred to this frame the Euler angles of the frame whose axes are normal to the planes of the canals are: $\psi = 0$, $\theta = 30°$, and $\varphi = 45°$ (Fig. 3). The equations of the transformation $\vec{\omega}\,(x, y, z) \rightarrow \vec{\omega}\,(\xi, \eta, \zeta)$ are given by Osgood (**29**) and also by Summers *et al.* (**41**). It is conceivable, of course, that motion out of the plane of a canal may nevertheless be capable of deflecting the cupula and thus of stimulating. In such a case, rotation in the plane of a canal would evoke an afferent frequency not only in the fibers from that canal, but in those of the other canals as well, and the central effect would be the resultant of all three effects. Such results have been reported (**1**, **16**, **21**). It seems to the author, however, that these are adequately explained by noting that because of the geometric disposition of the canals in the (human) head it is possible to choose an axis of rotation such that only the two horizontal canals will be stimulated, but it is impossible to choose an axis such that only the two anterior or the two posterior canals will be stimulated. Another conceivable difficulty would be the possibility that canals may interact mechanically with one another. This, however, does not seem likely from Lorente de No's result (cited by Spiegel and Sommers, **37**), that endolymph flow in one canal does not produce endolymph flow in the other two canals.

In closing this section at least brief mention should be made of stimulation thresholds. We have said above that it seems reasonable to take σ as the rate at which the central nervous system is being stimulated by the peripheral receptors. Assuming that there is some criterion by which one may decide that the central nervous system has responded, e.g., the onset of nystagmus, one may inquire as to what the smallest σ is which will just elicit the response. This is the rheobasic value of σ. A purely formal but convenient picture which has developed in physiology is that the stimulation builds up in the responding system some sort of excitatory state; at the same time this state is subject to dissipation by independent means. If in this offset the excitatory state ever reaches or exceeds a certain threshold value, however, it unstabilizes some hypothetical configuration and we have the response. In these terms it is easy to "explain" why any moderate rate of stimulation requires some time (so-called "latency") before it is able to evoke the response. The rheobasic value of σ (what is apparently referred to in psychological literature as the "threshold" value) has then an infinite latency. Carrying these concepts over to the motion-sickness problem, it is evident that whether or not the σ corresponding to a particular ship motion exceeds the threshold value for effecting sickness is a question of great importance. That this is not so for the σ's resulting from ship rotations seems to be rather widely accepted (e.g., by Sjoberg (**34**) and by McNally and Stuart (**21**); see also Morton *et al.* (**26**)). These conclusions should be considered with considerable caution, because in at least one instance (**9**) the various angular accelerations were not at all long enough to establish a true threshold.[9] Mowrer's (**27**) careful study was on a different species (pigeon) and may not be directly applicable to man; however, he found a nystagmus threshold (.79° per sec.^{2}) much lower than the average angular accelerations encountered on many ships.[10] In this connection it should also be noted that Spiegel *et al.* (**36**) have cited important evidence to show that the nystagmic threshold may be above that of responses of vegetative components of the autonomic nervous system associated with illness. It has also been shown (Henry, cited in reference **36**, and also Spiegel, **35**), that a typical partial symptom of motion sickness, the labyrinthine vasomotor reaction, may still be

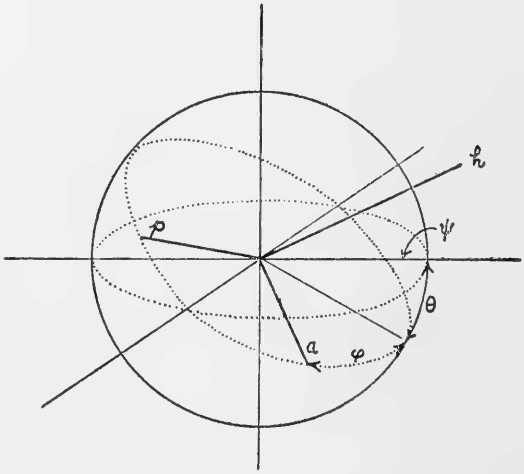

Fig. 3. Orientation of the left semicircular canals viewed from the left side. According to Summers et al. (23), the Euler angles ψ, θ, φ, have values 0°, 30°, and 45° respectively. The conventional axes, ξ, η, ζ, have been marked a, p, h, to indicate that they are normal to the planes of the anterior, posterior, and horizontal canals respectively.

[9] It should be noted parenthetically that the "thresholds to constant angular velocity" determined by Dodge were actually thresholds to unknown accelerations at the onset of the motion, for as we have seen in equation (8) a turning torque on the cupula acts only at the beginning of such motions.

[10] See, however, the summary tabulation of threshold determinations (**20**).

produced by rotating guinea pigs when the otolithic apparatus has been paralyzed by centrifugation. Previously, T. G. Brown (cited in **21**) had shown that the respiratory changes evoked by up and down movements were still present after removing both otoliths.

Studies Correlating Incidence of Motion Sickness with Physical Characteristics of the Imposed Motion

Studies falling under the heading of this section have apparently never been made at sea; therefore, we shall be concerned here only with laboratory experimental work. In view of the absence of data on the motion of vessels (see above), it is not surprising to find that the devices to induce sickness constructed during the war developed motions resembling ship or aircraft motion only in a general way. It seems also to be true that in the majority of cases such devices (e.g., hand-pushed swings) moved in a very complicated and not always reproducible fashion. One must therefore keep in mind the uncertainty surrounding the physical implications of the results.

Among the more definitive findings of wartime research are those having to do with the relation between head position and incidence of sickness on the swing. Manning and Stewart (**17**), using large numbers (70–100) of subjects per position, demonstrated the considerable superiority of the supine position in avoiding sickness on the swing. In their experiments the sitting position and the prone position both were associated with a comparatively high incidence. This was also the essential result of McIntyre (**19**). In a related study, Howlett and Brett (**13**) concur with the general finding that when the meatal-canthal line is parallel to the radius of the swing the incidence is least, and that when it is normal the incidence is greatest (however, these authors draw additional conclusions to those mentioned presently). Thus, for whatever accelerations the swing imposes, there appears to be established a directional effect. It is a difficult and somewhat harder problem to discover the time course of the accelerations, and a harder problem still to decide which receptors are involved. If a subject is swung with his head in some constant relation to the swing structure, as when held by a head rest (**6**), then the most natural way to resolve the accelerations is into radial and tangential components. An interesting attempt to do this has been made by Cipriani (**7**), on the assumption that the swing is an isolated conservative system.[11] He finds that the experimental conditions used with the Canadian swing are a period of 2 sec. and a radial acceleration maximum of about 0.7 g. The maximum tangential acceleration exclusive of the unknown contribution by the push is of the order[12] of 0.64 *g*. Cipriani regards the larger tangential acceleration as ineffective "since it affects equally the skull and its mobile contents." With this conclusion the present author disagrees, since the inertial terms of equation (3) depend precisely on the tangential acceleration. The possible effect of angular acceleration in their experiments was seemingly overlooked by the Canadian authors. Apparently on the assumption that only a radial acceleration operates effectively in the swing, Howlett and Brett have taken the aforementioned relation between the meatal-canthal line and incidence to suggest that the utricles are the receptors involved, since the plane of their maculae includes the line. (It is of course also true that the fixing of this line fixes other structures, for example, the vertical canals, in the correspondingly susceptible position.) These authors find further that under their hypothesis the utricular macula would have to be stimulated more when the otolith is pressed into

[11] Actually, of course, it is neither isolated (because energy is constantly imparted to it by external pushing) nor conservative (because there is friction).

[12] This is a rather uncertain calculation because it involves formulae which assume that the oscillations are small, and these are not.

it than when it is pulled away from it. It will be recalled that this is the reverse of Adrian's conclusion.

Of the many recent large-scale studies summarizing observations on incidence as a function of frequency and magnitude of stimulation, etc., the author has found it possible to draw physical inferences about the receptor system in only one case, viz., the important series of investigations by Wendt and his associates (**2–5** inclusive). In these studies human subjects (Naval

$$\text{S. I.} = \frac{2 \times \begin{matrix}\text{(number of)}\\ \text{(vomit cases)}\end{matrix} + 1 \times \begin{matrix}\text{(number of)}\\ \text{(nausea cases)}\end{matrix}}{\text{Total number of cases}} \times 100.$$

Statistical questions of experimental design and interpretation were carefully considered (**2**). In the following discussion, in addition to symbolism previously introduced, α will denote the acceleration level of an individual pulse, in multiples of g[13], and β the duration of an individual pulse, in seconds.

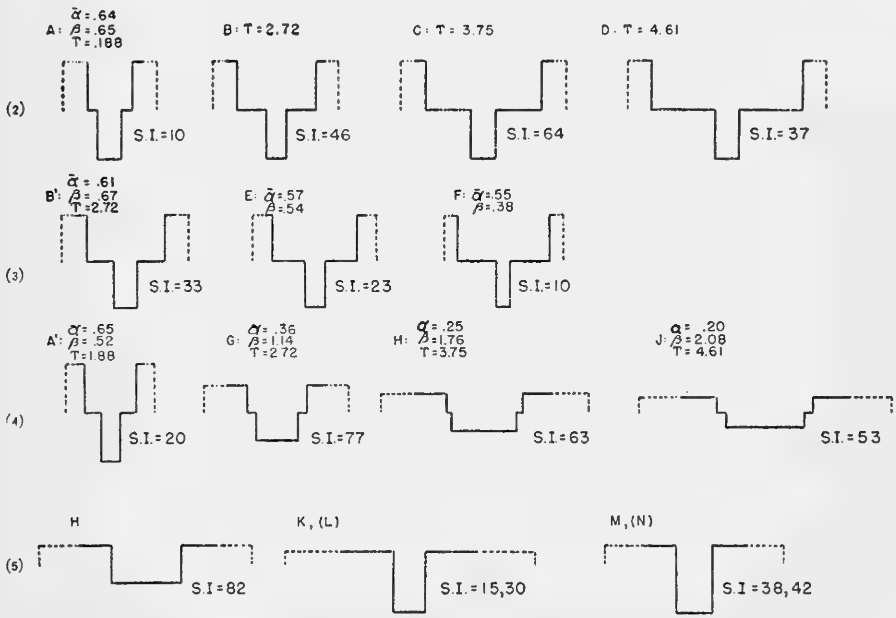

Fig. 4. Acceleration wave forms used by Wendt et al. The capital letters, A, B, etc., are the code designations used by Wendt; the numbers enclosed in parentheses refer to the bibliography of the present paper; the remaining symbolism is as in the text. In the case of asymmetric waves, the two sickness indices refer to results obtained with the wave as drawn, and with its inverse.

cadets and student officers) were placed in an elevator-like device, seated with ear eye line horizontal and blindfolded. The vertical acceleration imposed had an approximately square wave form, and features of the wave form were varied experimentally in a manner to be described below. Subjects were exposed to the motion for 20 minutes or until they vomited; of groups of about 100 subjects some 25% generally displayed either nausea or vomiting. The score employed was the "sickness index,"

Wendt's composite results are summarized graphically in Fig. 4, where the acceleration wave form appears together with the sickness index which it evoked. His general thesis in these studies appears to have been that the time characteristics of the forcing func-

[13] In some experiments α was not specified, and the value has then been calculated by the reviewer from the given changes in velocity and the "duration of the acceleration phase"; obviously such average values may be in error, and are always marked with a bar, thus, $\bar{\alpha}$.

tion are fully as important as the maximum amplitude developed during the cycle. In his final paper (**5**) dealing with physical characteristics Wendt concludes that the incidence of sickness can be described in terms of the response of "a resonant mechanical system with a natural period of 3.75 to 2.72 sec." He expresses the view, however, that such a system is not known (in the anatomical sense), and considers the vestibular system, for example, as being of "a very short period, heavily damped." It thus appears that at least in the broad sense Wendt was led by his experiments to form a model of the responsive system rather similar to the model of the receptors considered above. We shall here take issue with his second conclusion, which seems to exclude the vestibular apparatus as the responding system, but this is more a matter of detail. Aside from the question of the identity of the responding system, we shall suggest the possible general usefulness of the quantity, σ, in the analysis of experiments of the sort undertaken by Wendt.

It must be noted in connection with Wendt's experiments that, although they represent one of the best efforts in the field, the criterion—i.e., "sickness"—used to assess the nauseating power of the wave is an intrinsically unreliable one. This is evident from the varying results obtained with the same wave form (e.g., compare waves A and A′, and H and H′ in Fig. 4). The author is not prepared to offer a better suggestion, particularly because from the military operational point of view it is "sickness" which is the real criterion; nonetheless it is true that for purposes of physical evaluation the sickness index has to be regarded as a crude measure, and its variations are only rough indications of the response. A second point of caution to be observed in connection with the experiments of Fig. 4 is that the relation between incidence and the time of exposure to the wave has not been studied,[14] and one does not know where in the dose-response curve the 20-minute exposure lies.

Assuming, in accordance with considerations of the previous section and with the hypothesis of Wendt, that the imposed acceleration activates a responsive system consisting of a linear vibrator (this can of course be generalized to a spectrum of vibrators), we obtain $\nu\ (t)$ from equation (3) as soon as $f(t)$ is specified. In Wendt's case $f(t)$ represents a "square wave"; such waves can be given analytic expression by means of a Fourier series. When the proper series is substituted into (3) the resulting equation yields $\nu\ (t)$ also in series form. Putting this $\nu\ (t)$ into equation (5) gives finally the series for σ^2. This last series (for σ^2) converges rapidly when 2λ and κ^2 are of about the size calculated (**23**) for the canals; we can therefore base a qualitative discussion of the properties of σ on the square of the first term, which is proportional to

$$\frac{\alpha^2 \sin^2 (\pi\beta/T)}{[\kappa^2 - (2\pi/T)^2]^2 + 4\lambda^2(2\pi/T)^2} \qquad (6)$$

in the case of a symmetrical wave.[15] Some of the relevant features of expression (6) are the following: (a) Increasing the period from a minimum value (of 2β) to infinity, holding amplitude and pulse duration constant, reduces the numerator of σ^2 from a maximum to 0 as the sine^2 $(1/T)$, and also reduces the denominator as $(1/T^4) + (1/T^2)$. However, since at first the numerator decreases more slowly than the denominator the result is an increase in σ^2; eventually, the numerator tends to 0 while the denominator tends to κ^2, so σ^2 finally decreases to 0. This behavior can be paraphrased by saying that when one applies brief,

[14] It is the author's understanding that Professor Wendt has carried out experiments along these lines, but they have not yet appeared in the general literature and were not available to the author at the time of this writing.

[15] Analogous expressions for asymmetrical waves can be derived without difficulty.

closely-spaced pulses to a highly damped system, the time-integrated departure from rest is small because each pulse rapidly checks the effect of its predecessor; likewise, when the pulses are spaced too far apart, the same quantity is again small because it is computed on a per unit time basis. For intermediate spacings one obtains a maximum. It is interesting to note that when this variation of T holding α and β constant is carried out experimentally (**2**) the sickness index exhibits a behavior like the one just described for σ. (b) Decreasing pulse duration (β), holding amplitude and period constant decreases σ^2 as the sine^2; this is also in qualitative agreement with the observed behavior of the sickness index (**17**). (c) Increasing α, holding β and T constant should increase σ^2 as α^2—unfortunately this experiment was not performed by Wendt; in support of the result we have only the crude correlations such as were pointed out above that sickness is more prevalent on vessels subject to greater peak accelerations. Wendt's experiments with asymmetrical waves (**5**) could lead to information regarding the question as to whether the receptors involved react similarly in both directions, but the experiments cited here do not permit any certain conclusion because the pulse durations were not all equal.

A word should now be said regarding Wendt's conclusion that the vestibule is too "rapid" a system to account for what he calls the observed "natural frequency" (T = 3.75 to 2.72 sec.). The important experiment for this decision is **2,** Fig. 4. The sickness indices for increasing periods are 10, 46, 64, and 37. Taking the κ^2 and 2λ values previously calculated (**23**) from Lowenstein and Sand's data (**30**) for the semicircular canal, however, we obtain corresponding σ^2 values proportional to 5.73, 7.42, 8.28, and 8.48—that is to say, the period of maximum σ probably would have been *even longer* had Wendt used rays as subjects. Thus, whereas we are in no position to insist that the labyrinthine organs are the vibrators involved in this response, we do not feel that these experiments rule them out in any way. It also seems proper to suggest that the terms "natural frequency" and "resonance" be avoided except when they are used in their conventional physical meaning. If the labyrinthine receptors be overdamped vibrators, as has been suggested in the previous section, then they have no real resonance frequency or natural period. Regarding the behavior of σ^2 as a function of period or frequency in the experiment just discussed, it is evident that there is a critical frequency, ω_0, at which σ^2 or σ is a maximum, and should these measures of stimulation prove to be useful, it may be convenient to coin a word for said frequency. For the moment, however, a more important incidental matter is to note that for pulses of short duration $\omega_0^2 \cong \kappa^2$. This follows from solving $\partial\sigma^2/\partial\omega = 0$, assuming that $\beta/2$ is small enough so that $\sin \omega(\beta/2) \cong (\beta/2)\omega$. In principle, at least, this relation provides for an indirect experimental measurement of one of the characteristics of the vibrator, κ^2. Having measured this, the other characteristic, λ, could be obtained from the fact that the smaller[16] frequency, $\omega_{1/k}$, for which σ^2 is $(1/k)$th of σ_0, is related to ω_0 by,

$$\lambda = \frac{1}{\sqrt{k-1}}\left[\frac{\omega_0^2 - \omega_{1/k}^2}{2\omega_{1/k}}\right].$$

We have said above that Wendt and others have thought that the stimulating power of an imposed wave resides in its time characteristics as well as in its peak accelerations. On structural and analytic grounds this view seems amply justified, as we have tried to show. It also receives indirect support from considering the high accelerations which some experimentalists build into their devices in order to match the

[16] There is, of course, also a larger frequency, but this is not to be used in the equation for λ.

effectiveness of ship motion, although the latter is characterized by comparatively low accelerations. Such disparities lead one to speculate on the possibility that the time-course of the ship-imposed motion may have etiological consequences beyond the effects of labyrinthine stimulation, e.g., the establishment of an asynchrony between labyrinthine and other receptors (see above), and the possible handicap which irregular oscillations, as in a storm, may place on adaptive processes.[17]

Suggestions for Future Research

Although the gaps in the physical knowledge of the motion sickness phenomenon have probably been evident enough from the foregoing pages, it may be worth while to reiterate them here in more organized fashion.

It seems quite evident that for the motion-sickness problem and for other research problems as well—e.g., those of naval architecture, amphibious tactics, etc.—a rather extensive survey of the motions of various types of vessels in various seaways would be eminently worth while, and not too difficult or costly to make. Such data should include continuous records of roll, pitch, and scend, and it should be harmonically analyzed in the manner of Deacon (**8**). The correlation of such motions with sea conditions[18] assumes special importance in view of the recent advances in the forecasting of sea conditions near shores (**28**). Any quantitative knowledge of ship motion should find its way into the design of experimental sickness-inducing machines, not only in the interest of studying the effects of periodic stimulation, but also in the investigation of the "randomness" and phase-difference effects cited above as possible etiological factors.

[17] See Abels, H., cited in referencs (**20**).

[18] "Average" sea conditions have, of course, been investigated (see, for instance, the tabulation of H. Johow, cited in the review by McEachern, Morton, and Lehman, **18**).

Experiments of the sort carried out by Steinhausen, Ross, Lowenstein and Sand, and Adrian require great technical skill, but they appear to be among the most satisfactory ones for the purpose of elucidating receptor mechanics and neural response. On such preparations, for example, it would be very profitable to study the effect of oscillatory motion (which has, hereto, only been calculated theoretically), and to settle the important issue as to whether a canal can be stimulated by purely linear motion. The Adrian technique might be reapplied to extend our knowledge of the correlation between motion and the electrical response of the otolithic apparatus, among other things to learn whether response is two- or unidirectional. On the theoretical side, it appears worth while to formulate the motion of linear vibrators (of the otolithic and cupular type), disposed as they would be in the human head of an individual aboard a ship, and to calculate the σ-values for the conditions of the ship motion. This would require only a simple extension of the considerations of the second section above. The author is not qualified to venture opinions as to what particular approaches would be most profitable in studying central nervous phenomena, but only to acknowledge the great obstacle to quantitative description posed by present day ignorance of such phenomena.

The question of head position and susceptibility is an important practical one. Aside from its connection with receptor mechanics it might be worth while to study it in the course of specific operations, such as short transits in small landing craft (**42**), or at certain stations. Not necessarily the most immediately practical ones, but certainly the most basic studies in experimental production of seasickness are those of the type undertaken by Wendt. The author would venture to guess that the interpretation of such experiments will be easier with the aid of some theoretically-based analysis of the kind suggested here, although the specific one suggested here might be altogether too crude. The use of "square wave" acceleration pulses requires rather elaborate apparatus and a full-fledged harmonic analysis; in view of these complexities it might be worth while to consider the use of pure sinusoidal motion. If such has not already been done, a careful study of exposure time versus incidence would be in order, and in this connection it might be convenient to use in the analysis existing formal theories of neural behavior. A rather forlorn hope that someone may hit upon a superior criterion of incidence of sickness than hereto used would seem to be an appropriate closing recommendation.

Acknowledgement

The author is indebted to Professor Donald Lindsley and Dr. Herbert Landahl for suggestions incorporated in the manuscript, to Dr. James Birren, for valuable editing, and to Mr. Hugh Flynn for aid in obtaining many of the references here cited.

References

1. Adrian, E. D. Discharges from vestibular receptors in the cat. *J. Physiol.*, 1943, **101,** 389–407.
2. Alexander, S. J., Cotzin, M., Hill, C. J., Ricciuti, E. A., & Wendt, G. R. Wesleyan University studies on motion sickness. I. The effects of variation of time intervals between accelerations upon sickness rates. *J. Psychol.*, 1945, **19,** 49–62.
3. Alexander, S. J., Cotzin, M., Hill, C. J., Ricciuti, E. A., & Wendt, G. R. Wesleyan University studies on motion sickness. II. A second approach to the problem of the effects of variation of time intervals between accelerations upon sickness rates. *J. Psychol.*, 1945, **19,** 63–68.
4. Alexander, S. J., Cotzin, M., Hill, C. J., Ricciuti, E. A., & Wendt, G. R. Wesleyan University studies on motion sickness. III. The effects of various accelerations upon sickness rates. *J. Psychol.*, 1945, **20,** 3–8.
5. Alexander, S. J., Cotzin, M., Hill, C. J., Ricciuti, E. A., & Wendt, G. R. Wesleyan University studies on motion sickness. IV. The effects of waves containing two acceleration levels upon motion sickness. *J. Psychol.*, 1945, **20,** 9–18.
6. Cipriani, A. Adjustable head rest for swings. *National Research Council of Canada*, 1943, C4032.
7. Cipriani, A. An analysis of the forces encountered on the simple swing used in the study of motion sickness. National Research Council of Canada, 1942, C2246.
8. Deacon, G. E. R. Simultaneous frequency analyses of waves and ship movement. Admiralty Research Laboratory, 103.40/R. 1/W., 23 October, 1945.
9. Dodge, R. Thresholds of rotation. *J. exp. Psychol.*, 1923, **6,** 107–137.
10. Dohlman, G. Die Rolle der Perilymphe bei den vestibularen Reaktionen. *Arch. Ohren-, Nasen, Kehl. Heilkunde*, 1941, **150,** 25–30.
11. Dohlman, G. Investigations in the function of the semicircular canals. *Acta Oto-Laryngol.*, 1944, Suppl. **51,** 211–219.
12. Hasegawa, T. Über die labyrintharan Augenbewegungen bei Änderung der Kopflage und bei Progressivbewegungen. *Acta Oto-Laryngol.*, 1940, **28,** 593–600.
13. Howlett, J. G., & Brett, J. R. A speculation on the mechanism of utricular response to stimulation in motion sickness. *National Research Council of Canada*, 1943, C2509.
14. Jonason, I., Khylstedt, S., & Nylen, C. O. Tierversuche mit beschleunigter Rotation. *Acta Oto-Laryngol.*, 1940, **28,** 327–339.
15. Jones, H. M. The function of the labyrinth. *J. Laryngol. & Otol.*, 1945, **60,** 425–440.
16. Lowenstein, O., & Sand, A. The mechanism of the semicircular canal. A study of the responses of single fiber preparations to angular accelerations and to rotation at constant speed. *Proc. Roy. Soc. London*, 1940, **129B,** 256–275.
17. Manning, G. W., & Stewart, W. G. The effect of position on the incidence of swing sickness. *National Research Council of Canada*, 1942, C2426.
18. McEachern, D., Morton, G., & Lehman, P. Seasickness and other forms of motion sickness. *War Med.*, 1942, **2,** 410–428.
19. McIntyre, A. K. Effect of head position on susceptibility to motion sickness. Committee on Flying Personnel Research, RAAF, 1943, Rep. FR 86.
20. McNally, W. V. The otoliths and the part they play in man. *Laryngosc.*, 1944, **54,** 304–323.
21. McNally, W. F., & Stuart, E. A. Physiology of the labyrinth reviewed in relation to sea-

sickness and other forms of motion sickness. *War Med.*, 1942, **2**, 683–771.

22. Minkorski, N. *Introduction to non-linear mechanics. I. Topological methods of non-linear mechanics.* The David W. Taylor Model Basin, Report 534, 1944.
23. Morales, M. F. Asynchrony of labyrinthine receptors as a physical factor in motion sickness. *Bull. Math. Biophys.*, 1946, **8**, 147–157.
24. Morales, M. F., & Birren, J. E. Results appearing in a Nav. Med. Res. Inst. report classified for other reasons.
25. Morgan, R., Summers, R. D., & Reimann, S. P. Effects of various types of motion on differences in hydrostatic pressure between ends of a semicircular canal. *Arch. Otolaryngol.*, 1942, **36**, 691–703.
26. Morton, G., Cipriani, A., & McEachern, D. Mechanism of motion sickness. *Arch. Neurol. & Psychiat.*, 1947, **57**, 58–70.
27. Mowrer, O. H. The nystagmic response of the pigeon to constant angular acceleration at liminal and supraliminal intensities. *J. comp. Psychol.*, 1935, **19**, 177–193.
28. New York Academy of Sciences, Conference on Ocean Surface Waves, March 18–19, 1948. (Forthcoming publication)
29. Osgood, W. F. *Mechanics*, N. Y., Macmillan, 1937.
30. Ross, D. A., Electrical studies on the frog's labyrinth. *J. Physiol.*, 1936, **86**, 117–146.
31. Russell, H. E., & Chapman, L. R. *Principles of naval architecture.* II. N. Y. Soc. Nav. Arch. & Mar. Eng., 1939.
32. Schmaltz, G. The physical phenomena occurring in the semicircular canals during rotatory and thermic stimulation. *Proc. Roy. Soc. Med.*, 1931, **25**, 359–381.
33. Schmaltz, G. Versuche zu einer Theorie des Erregungsvorganges im Ohrlabyrinth. *Pflüger's Archiv*, 1925, **207**, 125–128.
34. Sjoberg, A. Experimentelle Studien über den Auslösungs Mechanismus der Seekrankheit. *Acta Oto-Laryngol.*, 1931, Suppl. 14.
35. Spiegel, E. A. Effect of labyrinthine reflexes on the vegetative nervous system. *Arch. Otolaryngol.*, 1946, **44**, 61–72.
36. Spiegel, E. A., Oppenheimer, M. J., Henny, G. C., & Wycis, H. T. Experimental production of motion sickness. *War Med.*, 1944, **6**, 283–290.
37. Spiegel, E. A., & Sommers, I. Vestibular mechanisms. In: *Medical Physics*, Chicago: Year Book Publishers, 1944.
38. Steinhausen, W. Über den Nachweis der Bewegung der Cupula in der intakten Bogengangsampulle des Labyrinths bei der natürlichen rotatorischen und calorischen Reizung. *Pflüger's Arch.*, 1931, **228**, 322-328.
39. Steinheusen, S. Über die Beobachtung der Cupula in den Bogengangsampullen des Labyrinths des lebenden Hechts. *Pflüger's Arch.*, 1933, **232**, 500–512.
40. Steinhausen, W. Über die durch Otolithen ausgelösten Kräfte. *Arch. ges. Physiol.*, 1935, **235**, 538–544.
41. Summers, R. D., Morgan, R., & Reimann, S. P. The semicircular canals as a device for vectorial resolution. *Arch. Otolaryngol.*, 1943, **37**, 219–237.
42. Tyler, D. B. The influence of a placebo, body position, and medication on motion-sickness. *Amer. J. Physiol.*, 1946, **146**, 458–466.

PART VII.

EMOTIONAL STABILITY AND ADJUSTMENT

Chapter 20

PSYCHOPHYSIOLOGY OF STRESS

CHESTER W. DARROW and CHARLES E. HENRY*

Institute for Juvenile Research, Chicago

Introduction

The pride of the service is that its men are tough, that only men who can "take it" are encouraged to remain, that the morale is of the best, that a man would be a submariner in preference to belonging to any other service. Yet, with the pride of officers and men in their dangerous and demanding service and with the well-schooled contempt for an easy berth or soft detail, there must be no needless disregard of human comfort or incidental wastage of highly trained and expensive human capacity. There must be no toughening at the sacrifice of efficiency. The unavoidable rigors of life on a submarine patrol subject men to conditions of severe psychological and physiological stress and demonstrate an endurance which bears testimony to the remarkable capacity of the human organism for adaptation, for compensation, and for acclimatization to conditions.

I. Analysis of Physiological Factors in Stress

Stress is defined in the Standard Dictionary as "That of which a force and its reaction are opposite aspects; a force together with its reaction, as in tension, compression, or torsion." The definition is particularly applicable to physiological and psychophysiological stress because of the emphasis upon *reaction* as an aspect of stress. Reactions to be considered here will be, first, physiologically determined adjustments to environmental conditions, often referred to as "homeostasis" (**13**), and second, psychologically determined and excessive or disorganized physiological changes which go beyond the requirements for homeostasis and which will here be referred to as "emotion." Physiological activity of both types may be sensed to varying degrees and may contribute to the background of awareness or consciousness, and, in relation to the background of experience, may be termed "affect" (**15**). Stresses prominent in life aboard a submarine may be classified conveniently under four general heads according to the characteristics of the initiating reaction.

A. Stresses Involving Severe Physiological Response to Environmental Conditions. These include stresses due to conditions of habitability, O_2 deficiency, CO_2 excess, temperature, humidity, vibration, pressure and even odors—conditions to which the organism responds automatically, physiologically, and reflexly. In this type of stress the psychological and emotional reaction of the subject occurs secondarily as a consequence of one's observation of his own changed physiological state. For example, a man's first awareness of depleted oxygen or excessive CO_2 may be the observation of his own increased breathing. He may refer to "difficulty in getting his breath," and he may become emotionally perturbed by the observation. Temperature and humidity in the same manner may be reacted to physiologically by cutaneous dilatation and sweating, and subsequently the increase may be observed and sensed as discomfort. Such stresses may thus involve two types of physiological reaction, (a) homeostatic, automatic, unconscious adjustment of the organism to the environment, and (b) the emotional (in-

* Now at Institute of Living, Hartford, Conn.

cluding autonomic) responses to sensations, with awareness or "consciousness" of the presence of these physiological adjustments. The initial physiological change and the emotional reaction to the change may in turn have secondary effects upon the central nervous system. All may alter the background of awareness or consciousness, and, when they have acquired "*value*" for the individual on the basis of association and "conditioning," may contribute to his "affective state."

B. Stresses Developed in Performance of Duty. A second type of stress is that due to demands placed on the performance of the organism. These are the stresses attributable to the rigors of performance of "duty," such as effort, fatigue, hunger, loss of sleep, etc. Emotional response may here be elicited on recognition of the demands of the situation. Further, different emotional responses may occur subsequently as the seaman becomes aware of his inadequacy or waning mental and physical ability to meet demands. Again, both the physiological consequences of the activity involved in performance and the emotional reaction to those effects may secondarily affect the central nervous system.

C. Emotion under Stress. A third type of stress is that initially involving a psychological response, for example as in depth bomb attack, with perception of the situation and the recognition of danger. Again, the physiological consequence of the emotion may secondarily feed back and affect the central nervous system. The fact that the perception of a dangerous situation may initiate processes in the brain which may discharge peripherally to increase heart rate, raise blood pressure, produce adrenaline, etc., has for years been the subject matter for those studying the "expressive" aspects of emotional response. The possibility that the autonomic, humoral and other changes associated with emotion may "feed back" and affect the vascular, chemical, and neuronal conditions of the brain has received less consideration. Of the possible importance of these "feedback" effects in accounting for the "functional" effects of emotion we shall speak in Section II C.

In the psychological reactions to the perception of danger the usefulness of the organic response may not always be obvious, but its physiological basis is readily understood. The increased muscle tonus, the increased heart rate, the increased blood pressure, the increased sweating, the adrenaline secretion, associated hyperglycemia, changes in the pituitary, and many of the other physiological effects which may occur in response to a purely perceptual recognition of danger are reactions which would be useful *if the submariner were engaged in severe bodily exertion* to meet the threatening situation. These are the changes which, if they occurred in support of the actual doing of work, would be regarded purely as preparatory or facilitative. That these changes occur in a perceptual situation in which the conditions do not require or permit violent expenditure of energy, and that they may, in the absence of an overt function, become disorganized and inappropriate, is apparently the only reason for labeling them "emotional" activities rather than merely homeostatic adjustments. The unique feature distinguishing emotion from other activity or "work" here lies in the fact that organic changes which normally would be organized to facilitate performance are present but that the activity or "work" is lacking. We do not incline to the view of James and Lange that a man fears a bear because he runs away from it. He experiences fear, rather, we believe, because he is producing organic physiological changes suitable for fleeing but in the circumstances is hindered effectually from doing so, and the organic changes, in consequence, are excessive, inappropriate, and may be disorganized. The hunted human animal lying in his bunk, walking without shoes, or talking in whispers while a stalking subchaser "pings" at his boat from above, may experience all the blood pressure, cardiac, vasomotor and gastrointestinal changes he might have if he were permitted to run or to fight. Under the circumstances the organic changes may be described primarily as emotional, whereas if the same changes were associated with running or fighting, the organic activities would be considered physiologically adaptive, and would be regarded as "emotional" only secondarily and to such extent as they were in excess of and inappropriate to the physiological need. More important, if possible, is the fact that these organic changes—increase of blood pressure, etc., occurring inappropriately, in excess, and without metabolic demand—may in turn set up extended sequences of organic compensatory changes or "side reactions" which may be observed later as disturbances of digestion and of central nervous function persisting for hours and days following the experience. In physiology as in pharmacology we may often have to consider secondary as well as primary sites of action. A gastrointestinal upset may, for example, represent a *compensatory* "side effect" of the sympathicoadrenal mobilization consequent to a disagreement with one's superior officer. The nausea and anorexia sometimes reported days following depth bombing may be such an effect. Reactions of this kind may on occasion characterize operational stress, combat fatigue, etc., and accordingly they deserve serious consideration.

D. Conditions of Physiologic Impairment Accessory to and Aggravating Effects of Stress. Of extremely great importance for present consideration must also be a number of conditions not directly or strictly resulting in stress. These include certain conditions of habitability and other factors for which there is no direct physiological compensation. These are conditions which, persisting over a period, may become debilitating. Among such conditions are the loss of physical tone due to deficiency of dietary factors, lack of vitamines A, B, and C in particular, exhaustion, irregularity of meals, impairment of autonomic balance, excessive temperature, humidity, skin irritation, loss of sleep, and adrenocortical deficiency. On the psychological side are boredom, loss of *esprit de corps*, lack of "morale," and lack of interest in performance of a mission. This latter state of mind has been noted to result on long and routine patrols, and has been observed to improve following successful engagement of the enemy, with effects even on the physical states of the men. With this in mind the possible role of the "state of consciousness" will be considered as a factor in response to stress. (See Section IV.)

II. Evaluation of Conditions of Stress in Undersea Warfare

The physiological effects of the outstanding types of stress such as stresses due to oxygen lack, to extremes of temperature, to fatigue, to loss of sleep, and even to fear, have been extensively studied and the results summarized in the literature.[1] There is little point in reannotating an already extensive bibliography on these topics. A profitable consideration would:

a. Show the relation of these homeostatic and emergency functions to prevailing or subsequent emotional states.

b. Show how the organic, autonomic and other physiological conditions feed back into and affect the nervous system.

c. Show how rational evaluation of these conditions may provide better means of adjusting and selecting human capacity for the type of psychophysiological stress encountered in undersea warfare.

d. Evaluate methods of testing and selecting human capacity for the type of psychological stress encountered in undersea warfare.

e. Suggest desirable changes in submarines or their mode of operation.

[1] Cf. **1, 3, 5, 6, 7, 10, 11, 12, 13, 23, 30, 31, 33, 36, 45, 52, 64, 70, 76, 85, 86, 90, 93, 99, 102.**

A. Physiological Conditions

1. Oxygen and carbon dioxide

We shall not review the problem of habitability, O_2, CO_2, humidity, temperature, and related problems in detail, since these have been treated in other chapters. Our concern is primarily with the physiological mechanisms of adjustment—i.e., reactions by which the organism maintains the internal milieu. It is these mechanisms which determine the manner of adjustment to the important conditions of stress. The reactions of these mechanisms are of importance not merely because they provide homeostatic control in maintaining the internal environment. A knowledge of their operation may also throw light upon some of the coincidental or "side effects" that may occur secondarily to the main adjustment.

The oxygen and carbon dioxide in the respired air does not present the same problem to the submariner that it does to the aviator or mountain climber (**85**, **86**, **98**). Not only is the oxygen deficiency of high altitudes not associated with an increase of respired carbon dioxide as in the submarine, but in conditions of flight the increased rate and depth of breathing in response to deficient oxygen results in further depletion of CO_2 with consequences physiologically (**29**) which may be quite different from those in the submarine where the carbon dioxide of the respired air may go as high as 3% and oxygen may be reduced to 17% in a 20- to 30-hour period under the surface. In consequence, much of the valuable research which has built up an extensive knowledge regarding the reactions of the organism to oxygen deficiency is not applicable in a submarine, where depletion of oxygen is ordinarily attended by accumulation of CO_2. (Cf. Section V, 1.)

The submariner is relatively fortunate in that conditions giving rise to decreased O_2 lead to an increase of CO_2. An increase in CO_2 to the level of 2–4% in the inspired air has dramatic effects, postponing or preventing syncope and helping to maintain blood pressure even during hyperventilation. For example, breathing of 8% oxygen in a standing posture will produce a marked hyperventilation associated with a fall in blood pressure, increase of heart rate, and syncope in the average male subject. Addition of 3% CO_2 to the mixture will in many cases prevent the fall in blood pressure and other consequences (**48, 50**).

CO_2 also helps to maintain venous pressure via effects on the sympathetic system, and this provides a better return of blood to the heart with consequent support of blood pressure and maintenance of supply of blood to the brain. CO_2 decreases the alkalosis, shifting the O_2 dissociation curve and enhancing the ease with which the increased blood supply can unload O_2 in the body tissues (**96**). Particularly important is the effect of CO_2 on brain circulation. The powerful vasodilation effects of CO_2 on cerebral blood vessels markedly increases the supply of oxygenated blood to the brain (**47, 50, 51, 54, 92, 97**).

Many studies could be cited dealing with the details of these complex reactions constituting the sequence of events occurring in the body's adjustment to the stress of oxygen lack. In one Navy study (**37, 39**) low O_2 was used as a stress situation and changes in the electroencephalogram (EEG), interaction chronogram, adrenal cortical hormones, lymphocytes, pulmonary ventilation, heart rate, blood pressure, and certain psychological changes were studied on normal controls and on subjects with psychoneurosis and combat fatigue. Whereas 8 to 10% O_2 resulted in a clear shift of the distribution curve of occipital alpha frequency to the slow side, even slight amounts of CO_2 (0.3–0.5%) definitely reduced the slowing, whether measured by hand or with the Walter analyzer (**4**).

Earlier work (Gellhorn, **48, 50,** Gibbs, **53,** and others) using more severe stress (lower O_2) emphasized the importance of CO_2 for maintenance of normal brain function. As an extreme example (**53**), two subjects were able to maintain EEG stability and good mental contact for two minutes despite a very low O_2 supply (2%) in the presence of 5% CO_2. Furthermore, the intermediate stages of increased voltage and decreased frequency in the EEG were greatly reduced.

Whether one regards changes in blood or in brain as primary in these reactions, it should be remembered that peripheral neurophysiological events are also taking place as part of the total organismic response. There is increased sympathetic activity in anoxia with release of adrenaline to play its supporting role in circulatory changes. Chemoreceptors in carotid and aortic glomi serve to feed back and affect the nervous system, and pressure receptors in these areas reflexly assist in maintenance of heart rate and blood pressure. The effectiveness of the adjustments may be furthered by a sitting or reclining posture, reduced physical exertion and maintenance of adequate *blood sugar* (**29, 51, 67**). This latter has received extensive study, but it is merely mentioned here, since sugar *per se* is unlikely to be a critical factor in submarine life where food supplies may usually be depended upon to outlast O_2 supplies. Suffice it to say that glucose and O_2 play mutually supportive roles in brain metabolism and that normal and even supra-normal blood sugar values (130 mg%) may serve to compensate somewhat for reduced O_2 supply (**16, 29, 51, 63**).

In view of the variety of skilled motor tasks, sensory-motor adjustments, and decisions involving multiple operations affecting the entire submarine, it may be of interest to examine data secured by McFarland (**85**) with regard to the effects of anoxia on certain psychological tests. It was found that, on the average, there was significant deterioration at 14,000 feet (12% O_2) with respect to handwriting, choice reaction time, and ocular muscle balance. More complex tasks such as color naming, a transliteration code test, and a paired associates memory test showed impairment at 10,000 feet (14% O_2). There was considerable individual variability in all these tests with some individuals showing effects at higher O_2 values (cf. **70**). Rate of ascent and oxygen deprivation was an important factor, the effect being greater for anoxia of rapid onset. Under extreme conditions, of course, the danger of incurring the "bends," as in explosive decompression (**67**), introduces still another factor.

The psychological changes that parallel the physiological alterations due to hypoxia are equally important, although somewhat more difficult to detect in the earlier stages. With slowly developing anoxemia the changes are characteristically insidious of onset and not readily recognized by the individual. There is usually an initial feeling of well-being, perhaps similar to the effects of mild alcoholic intoxication, with feelings of power and self-satisfaction and decreased critical capacity (**3, 16, 48, 85**). With progressively greater anoxia there is impairment of judgment and emotional instability coupled with complaints of frontal headache and shortness of breath (**58**). Loss of critical capacity combined with muscular incoordination and general irritability may result in rough handling of equipment or an inability to carry out procedures necessary for safety and survival of the crew. There may be a fixity of ideas and perseveration of (sometimes erroneous) responses. There is both sensory and motor loss, particularly of visual functions, and impairment of memory. With severe or prolonged anoxia there may be nausea and vomiting, extreme weakness and eventually convulsions and cardiac syncope. In the terminal stages of anoxic anoxemia, blood pressure, heart rate and circulatory responses are extremely variable. Relationships between susceptibility to low O_2 and the presence of various "psychoneurotic" traits provides some basis for the use of low oxygen in various screening procedures to be mentioned later (**30**).

Some of the anoxic effects on sensory systems may be of particular interest to the submariner because of the importance of competent lookouts and watches. The same is true of effects on sensory-motor (reaction time) changes. Although many of these data have been obtained relative to flight problems where lower O_2 and CO_2 values prevail than in submarines, they are mentioned here because they constitute basic research some of which is applicable to the submarine. (See Section V, 1.)

Vision. There is evidence that O_2 lack and CO_2 excess exert their effects somewhere central of the photoreceptor systems (**3**). While the course and shape of the dark adaptation curve shows little change, anoxia decreases visual sensitivity by more than one-half at 13% O_2 (**51, 52, 85**). Hyperventilation and the blowing off of CO_2 add to this effect, further decreasing visual sensitivity. Hypoglycemia likewise increases the visual threshold. The latency of negative after images is also increased by anoxia (**52**).

That impairment of visual function may be cumulative and result from even mild degrees of anoxia is shown by recent University of Chicago studies (**16**) using the critical fusion frequency (cff) and the dynamic visual field (dvf) test as measures of performance (**58**). The dvf test is a contrast discrimination involving both central and peripheral fields. Of 20 men passing Air Corps standards exposed to low O_2 (pressure chamber) for five to six hours per day, six days per week for four to six weeks, 13 (65%) showed visual impairment even at 10,000 feet (14% O_2). It is of interest that this impairment took several weeks to develop, that it could be prevented by breathing pure O_2 for one hour daily, and that the individuals were *completely unaware* of their impairment. Recovery was delayed for several weeks and subsequent re-exposure resulted in more rapid deterioration. Tests of cff for intermittent light also showed a clear impairment at 10,000 feet.

It should be emphasized that the performance decrement on these tests was found at oxygen levels where intelligence, determined by the usual psychometric tests, was unaffected. The insidious onset of this relative blindness with the accompanying minor complaints of headache, blurred vision, drowsiness and lassitude, depression, restlessness, and even euphoria could easily result in undetected and potentially dangerous degree of incompetence. This would be particularly likely to occur when men "stuck it out" without complaining.

Consciously resolving not to "give up" and ask for relief introduces subtle changes in the inter-relationship between the individual and his job. He may become irritable and susceptible to emotional upsets over his errors (with perhaps a tendency to "cover up"—although normally he would realize the possible fatal consequences of so doing). Further changes correlated with his increased effort could lead to inappropriate psychophysiological effects, along with a severe conflict and frustration. How far such a sequence might progress would be a function of numerous variables including the stability and resources of the individual; his degree of fatigue and impairment; his confidence in the boat, its skipper, and his mates; and his evaluation of the factor of potential relief or removal.

Hearing. The effect of anoxia on auditory thresholds is similar to the effect on vision. The degree of loss is a function of both duration of exposure and degree of anoxia (Gellhorn, **46**); 10% O_2 for 15 to 30 minutes sometimes results in a hearing deficit for several hours. Similarly to its effect on vision, alkalosis induced by hyperventilation in room air also decreased auditory activity. Any impairment resulting from the chronic moderate-to-loud noise level of submarine operation would presumably be aggravated by low O_2.

Further complication, because of the interrelationship with both visual and auditory functions, may well come from chronic subacute vestibular stimulation. It is probable that the "noise" or wave motion spectrum on a submarine extends over a very wide range;

perhaps from a slow 20 second roll of the boat through the intermediate range of vibration and well up into auditory frequencies. It has been long realized that noise can be a potent distracting and interfering factor in many kinds of activities. While these subjects are treated separately in other chapters, we wish to emphasize the possible cumulative effects of these manifold and completely unavoidable stimuli, and to suggest that their effects may be something other than purely additive on a severely stressed organism.

A possible further corollary here may be body temperature, which tends to drop during anoxia, thus reducing metabolic demand and compensating somewhat for the diminished O_2 supply (**49, 74**).

2. Temperature and Humidity

Temperature control is, of course, only one facet of the regulation of energy exchange between the body and its environment. The submariner is homeothermic to a remarkable degree, even if his air-conditioned submarine is not. The close control of body temperature is a prerequisite for adequate functioning of other homeostatic systems, for without it, basic biochemical and metabolic reactions could not be maintained. The critical problem in submarines is usually one of heat dissipation; this may occur through radiation, conduction, and water vaporization. The latter is the most effective but is most critically affected by humidity.

The autonomic homeostatic mechanisms that maintain temperature control have been widely studied (**45, 64**). Experimentally placed lesions in the hypothalamic thermostat have revealed that there are separate but anatomically close regions that are sensitive to the temperature of the circulating blood. The more caudal hypothalamic region is activated by relative cold, and gives rise to shivering and muscle tension, vasoconstriction, piloerection, and adrenaline secretion—sympathetic reactions. The more rostral region responds to warmer blood with consequent vasodilatation and sweating. The range of change in blood flow in peripheral tissues is high—of the order of 100:1.

That both sympathico-adrenal and vago-insulin mechanisms are activated by exposure to both heat and cold and that it is only the relative dominance of one over the other that provides regulation appear well supported by Gellhorn (**52**). The predominating effect of excessive heat and humidity apparently involves both a mobilization of the vago-insulin or parasympathetic system and an inhibition of the sympathico-adrenal system. Vaso-dilatation, sweating, and increased respiration are the chief mechanisms of heat elimination. The result of dilatation and of hyperpnea may be a loss of venous pressure, and, presumably, by way of pressure receptors and inhibition of parasympathetic activity, a marked increased in heart rate. Increased rectal temperature, lowered blood pressure, tachycardia, and reduction of CO_2 by hyperpnea are conditions which have been found indicative of failing ability to adjust to high temperature (**1, 5, 7, 36, 71, 74, 99**). EEG studies have shown that these conditions may be associated with impairment of cerebral function.

The skin and lungs are the main lines of defense in temperature control. In the submarine, as in other environments, protection by clothing exerts an important influence on heat exchange. Except in extreme heat, nude men can work in higher temperature and higher humidity; e.g., at 90 degrees F. some evaporative cooling occurs even at 99% humidity, while with clothed men this limit is 89% (**99**). With elevated dry and wet bulb temperature extreme discomfort and sharply decreased working efficiency will result. Since muscular work greatly increases heat production in the body, the job of the planesman, for example, in silent operation with air conditioning and power controls off is a severely taxing one. Facilities to increase air movement would be helpful.

Capacity to carry on efficiently at high temperatures depends on more than general physical fitness. The capacity is highly specific and is acquired. It is a function of repeated exposures and of time (**1, 5**). The importance of training in anticipation of exposure to excessive temperature can hardly be over-estimated. The critical importance of abundance of drinking water and adequate replacement of salt has been experimentally demonstrated by Bean and Eichna (**5**). One is impressed by the rapid and dramatic effects of salt and water deficiency upon psychophysiological status. (See Section V, 2.)

The limiting conditions of habitability as shown by Eichna *et al.* (**36**) for various combat vehicles are dependent on wet bulb temperatures. They report that as the upper limit is approached a narrow range of wet bulb temperature differentiates the environment "where work is easy" from "where it is impossible." Theoretically the problem should be to obtain that combination of physiological activities which will make the best use of narrowing possibilities for heat dissipation into the environment. This may involve a complex balancing of conflicting physiological tendencies. Of course, practicable devices for reducing environmental temperature and humidity reduce by that much the demands on the physiological mechanisms. (See Section V, 3.)

3. Motion Sickness

It is not possible here to review the extensive research toward a means for control of seasickness (see chapters 18 and 19, on Motion Sickness). However, it is mentioned here as an example of what may be done by well directed investigation of a psychophysiological problem. In this case the control of autonomic functions, chiefly by the use of anticholinergic drugs, has provided an improved control of one psychophysiological condition. (See section V, 4.)

4. Odors

The intensity of the olfactory stimulation and the magnitude of the gastrointestinal responses to it are testified to by repeated reports from the submarines which have undergone prolonged subsurface duty. The intensity of the stimulation becomes obvious only when life below is interrupted by watch on deck. This offers convincing testimony to the effectiveness of olfactory fatigue and adaptation as protective mechanisms and contributory factors in habitability.

B. Stress of Duty

The physiological concomitants of activity, in performance of duty, span the gamut of human activity. They may range from those of inactivity and rest through the various degress of activity, mild, moderate, strong, severe, and exhausting. Attending psychological states may range from (a) somnolence, indifference or boredom through (b) alertness, interest and attention, (c) tension and anxiety, (d) intense effort, to (e) weariness, frustration and confusion. Although only the more severe degrees of these conditions are commonly identified with stress, we may not ignore the possibility of severe secondary psychological effects correlated with enforced unwelcome inactivity. The adverse effect of monotonous patrolling with no hits has been too frequently noted. (See Section V, 5.)

The evaluation and measurement of conditions of stress may seem to approach an impasse, as have studies of "work" and "fatigue," because of differences in definition and confusion as to the nature of what is measured (**3, 6, 8, 10, 31**). There may be confusion between decrement in quantity and decrement in quality of output. At one time we may be concerned with (a) physiological or organic effects, at another with (b) psychological or subjective effects, and at another with (c) behavioral effects and alterations of output and performance. These may be measured respectively (a) by physiological indicators such as metabolism, (b) by subjective criteria such as ratings of "satisfyingness" of work or feeling of fatigue, and (c) by measurements of performance or output of work. In the consideration of stress we are concerned with effects of all three types. Yet, if we may be guided by experience in attempts at the study of fatigue, we need expect no close correspondence among measures of these different aspects of stress even when obtained simultaneously on the same subject. This need not mean that the different measures are unreliable, but only that these different aspects of human function are not necessarily closely related and proportional.

The mobilization of the complex physiological organism for varying degrees of activity not too remotely resembles the wartime mobilization of a nation for an offensive. Not merely the striking arm at the front, but the entire energy and material producing resources of even remote parts, are marshalled to meet over-all demand. Activity in one part may support that in another—and no less may seriously interfere with that in another. This is strikingly illustrated by the rising patterns of tension which develop with increasing intensity of activity and motivation. The amount of work done may thereby first increase as the result of the greater effort, and later decrease (**40, 95**).

The ideal state of rest is probably one of minimal muscular tension or maximum relaxation (**68**). The ideal waking, alert, erect position, on the other hand, with shifting position of the head involves complex patterns of automatically maintained tensions or "tone" as have been so beautifully demonstrated by Magnus and de Kleijn (**80**) and reviewed by Fulton (**45, 64**). Such tensions involve not merely efferent neural discharge into the musculature, but also afferent proprioceptive and other discharge back into the central nervous system, thereby providing a circular self-perpetuating *background* (**22**) of activity. Upon this background, as by a finger painter, shifting concentrations of the "*postural substrate*" (**42**) sketch the pattern of response. The existence of a certain amount of prior tension or "postural substrate" in the system appears to be essential for effective adjustive activity.

Along with this elevation of postural tone there occurs a general mobilization of various preparatory and facilitative (**21**) organic processes, notable among which are changes in blood pressure and palmar sweating. The latter activity has proved of particular value as an index of muscular activity (**56, 100**), alertness, and general mobilization of the organism (**32**). It will be referred to later in discussion of test procedures.

But whereas increases of background tension may, within moderate limits, increase the proficiency of response and explain why various types of stimulation and voluntarily maintained muscular tension may sometimes improve performance (**9, 18, 41, 100**), a high level of tension in conditions of effort, conflict, and frustration may be reached beyond which further increments no longer provide further facilitation (**18**). This is possibly the point at which we begin to think of a situation as involving "stress." It is possibly somewhat beyond this point that we reach the level representing the "Plimsoll mark" (**43**) of the individual's capacity, beyond which the "load" should not further be increased. In the normal tonic state moderate shifts of tension apparently affect and facilitate the specific effectors involved in an activity. This may account for facilitation. With increasing tension and stress as indicated by increased intensity of activity or by increased exhaustion and fatigue of already active mechanisms it appears that the increase of "tension" with activity may affect not only the specific and appropriate musculature, but larger and less appropriate muscles and groups of muscles as well. The activity may appear to be performed with continued effectiveness, scores on performance may be unchanged or even improved, but efficiency in terms of work done per unit of energy consumed is greatly reduced (**6, 8, 31, 40**)—the effort is out of proportion to the performance. (See Section V, 6.) And as the stress of increasing effort rises, the spread of tension may increase until the activities involved are no longer even appropriate and may even interfere with one another (**8, 18, 40**). Accessory autonomic, glandular, and other normally supporting activities may be mobilized out of all proportion and even at times inappropriately to the specific task. Homeostatic readjustments of the organic factors in equilibrium may be initiated and may persist after the immediate occasion for stress has subsided (**2, 78, 101**). Even appropriate supporting organic changes, as for example the secretion of adrenalin, may have undesirable "side effects" or give rise to undesirable compensations. Thus, as the organic changes attending performance of duty become excessive or inappropriate they may by this fact acquire the qualification of an "emotional" response. And as they become effects secondarily and more remotely induced they may, as we shall try to show, provide symptoms of "functional" disorders (**34, 55, 82**).

C. Stress of Perceived Danger-Emotion

That the perception of external conditions constituting a threat to the safety or security of the individual may give rise to psychophysiological stress is a natural inference from the nature of the profound physiological changes which may be produced. But of the nature of the mechanisms by which perceptual patterns projected on the "mental screen" are identified and interpreted as carriers of friendly, neutral, or enemy cargo and, for the latter, translated into a firing signal setting off an appropriately directed reaction, our knowledge is limited. It is true that psychophysiological "conflict" in such conditions can often be recognized and demonstrated (Luria **79**). In an early paper one of the writers postulated central "conflict" among impulses as a condition involving a degree of "functional decortication" which might release subcortical mechanism from central control (**20**). While decortication is consistent with the automatic, involuntary and irrational, over-reactive, incoordinated, asynergic, dysmetric character of the typical emotional reaction, sometimes involving loss of acquired skills and sometimes subsequent amnesia for the event, the process is obviously not a simple one. Kennard (**69**) recognized the interrelation of emotional states, incoordination, and basal ganglionic dysfunction. Present evidence suggests that there is not only a central excitation process which may discharge peripheralward to produce autonomic changes, but that there are also effects on basal ganglia as well as hypothalamus, which may in turn react back upon the cortex (**27**). These latter will be considered in Section II, C.

That sensory stimulation effectively gives rise to cortical effects; that previous experience may "condition" or "sensitize" the organism toward specific types of stimuli (**2, 56, 66, 83**); and that by man's facility with symbols he may by suggestion and rationalization become sensitized to react to a wide range of mentally perceived hazards with which he never has had prior experience, are not matters for controversy. We nevertheless are profoundly ignorant of the neurophysiological mechanisms by which these effects take place.

The attempt to study emotion in the laboratory encounters difficulty in the inadequacy of the laboratory situation. There is always an element of artificiality, of unreality, of the absence of life and death consequences. Experimenters have tried to provide reality for their experiments by the use of painful stimuli such as heat or electroshocks (**41, 86, 89**); they have tried removal of postural support; they have employed "frustration"; and they have sought to suggest reality by resort to hypnosis. But when set beside the harrowing experience of depth bombing, for example, these may not seem too convincing. Even the most satisfactory laboratory experiment on emotion bears more weight when its conclusions may be supplemented by observations under stresses of real life conditions. (See Section V, 6.)

On board a submarine, on patrol or in combat, it is conceivable that the stimulating effects of perceived danger may be out of all proportion to anything produced in a laboratory, and they may be greater than those usually encountered in life situations. The limitation of possible action may magnify the perceptual effect. Evidence from the laboratory seems to indicate that limitation of opportunity for motor activity under stress may increase the emotional consequences of stimulation (**44**). Without such expression every sound of the pursuing enemy or by ones colleagues on board may be exaggerated. "The drawing of a glass of water may sound like Niagara." The confines of a submarine permitting but limited motion or adaptive activity might conceivably further aggravate the discrepancy between mobilization and actual expenditure of one's energies. Not only is attention not objectively directed, but there may also be a relative absence of overt expressive activity. Lack of opportunity for aggressive action has been reported bad for morale. Anger during depth bombing, for example, has been attributed to the necessity for "taking a beating without fighting back." The urge of the submariner to seek refuge in action and fight it out among the (for him) even greater dangers of surface warfare may represent such a need for action. One has the feeling that to improvise duties for these men (oxygen supply being adequate), to keep them intensely active and provide a nucleus of organization and a function for otherwise "free floating" emotional organic changes, would be doing them a service.

To the extent that reactions are excessive, disorganized, or inappropriate to the situation they have been designated "emotional." By definition such reactions have ceased to serve solely as preparatory and facilitative physiological changes. Their added value if any, must be a psychological one, an effect on the man himself or on his confreres. And reactions to the observation of one's own extreme physiological changes may also be classed as emotion. Whether such reactions occur on observing one's own homeostatic adjustments to conditions of habitability, his own behavior under stress of duty, or his own reactions to perception of dangers threatening from without, they by definition exceed the immediate physiological need. A person, for example, may become emotionally perturbed on observing his own cardiac changes in response to an injection of adrenalin in a laboratory experiment.

Feedback Effects of Emotion Upon the Central Nervous System

In the early study of emotion by the employment of "expressive" methods it was assumed that the analysis of such symptoms would provide clues as to what was going on within the organism. The peripheral changes recorded by "expressive methods" were considered as symptoms, and little consideration was given to the possibility that these organic changes might themselves have important effects upon the brain. With the development of EEG for recording of the brain activity electrically, and with the concomitant recording of autonomic activities by the polygraph (**26, 27, 60, 61, 76**), the effects of emotion or autonomic changes upon the central mechanisms have become demonstrable. Now that it is recognized that engine governors, thermo-regulations, self-stokers, self-feeding, self-piloting, self-aiming, self-guiding machines are possible only by application of the *feed-back* principle (**84, 91**), we have come to see examples of feedback in nature where its importance was never suspected before. It is realized that without such systems a coordinated, directed, regulated system such as the human body would be impossible.

What happens when the human regulatory systems are damaged may be observed in the dysmetrias, incoordinations, palsies, athetoses, and adiadochokineses on any neurological ward. What may happen in these systems under the stress of emotion has been pointed out by Kennard (**69**). The physiological changes observed in emotion may be no less of importance because of their afferent than because of their efferent or expressive manifestations.

Channels of such feedback effects are numerous (**26, 52, 84, 91, 93**). They include the whole range of psychophysiological activity, humoral, chemical autonomic, neurophysiological, and, we shall try to show, psychological. Humoral effects upon the brain involving particularly effects of adrenalin, pituitrin, insulin, sympathin, and acetylcholine probably take many forms, but only one of these, the most definite and well established, the effects upon the blood vessels, will be considered. The susceptibility of the cerebral blood vessels to changes in carbon dioxide has been well established (**24, 25, 48, 50, 52, 92**). The effects of hyperventilation with blowing off of carbon dioxide are readily demonstrable in susceptible persons by means of the electroencephalogram (**27, 29, 76**). As little as three percent carbon dioxide or a cerebral vasodilator such as glyceryl trinitrite may prevent the effect (**92**).

That the parasympathetic supply to the brain via the facial and greater superfical petrosal nerves may serve as a regulatory feed-back mechanism has been demonstrated (**17, 24**). Feed-back effects via sympathetic pathways are suggested by the work of Murphy and Gellhorn (**87, 88**).

That the entire sensory system, including the proprioceptive afferent systems, provides a mechanism controlling bodily responses is commonly recognized (**80**), but that these afferent processes serve as parts of an extensive integrated coordinating feedback mechanism maintaining tone and governing reaction needs emphasis. That they likewise contribute substance to that complex subjective state called "consciousness" will be considered in Section IV.

D. Stress as a Function of Chronic and Accumulated Impairment

Long submarine patrols where attendant dangers and discomforts are extended in time, and where the conditions of stress become more or less chronic, introduce another factor: impairment of the capacity of the organism to react to stress. These conditions are likely to result in true organic or physiological impairment of the type that can be demonstrated by microscopic and chemical methods. Cumulation of these effects may eventually lead to a condition such that the boat can not be competently manned in all of its functions. Under such conditions persistence of psychophysiological effects after the source of disturbance has been removed may be identified in various forms of "operational stress" or "combat fatigue" (**16, 34, 55, 82**).

1. Sleep

The capacity of the organism to function adequately despite wide variations in the amount and distribution of its usual sleep periods has received much clinical and laboratory study, although there is still some disagreement with respect to the amount of sleep that is necessary and desirable. The watch system of the Navy has always required a flexible and frequently interrupted schedule. Furthermore, in submarine duty the usual light-dark diurnal cycle may be missing or reversed for weeks on end. There is some evidence that younger individuals adapt more readily to imposed changes in their wake-sleep rhythm (**73**).

Even before conditions of extreme sleeplack are encountered there are performance changes that are of vital interest to a submariner. Rather surprisingly, these changes are not marked for such tests as steadiness, tapping, static ataxia and visual acuity, or for such measures as blood pressure, heart rate, temperature, and basal metabolism (**3**). There is increased sensitivity to pain and, more important, loss of memory, hallucinations and delusions, semi-dreaming states, depersonalization, double vision, and consistent increase of irritability. Experiments, which have run up to 100 hours of sleep deprivation, also showed that in well motivated subjects fatigue and boredom were not usually marked, although irritability is aggravated (**3, 73**). Diurnal rhythms tend to be preserved. Activity and moving about may be necessary to keep from falling asleep. (See Section V, 7.) Effects of prolonged sleep loss have been reviewed elsewhere (**73**) and are considered in detail in another section of this book (Chapter 15).

2. Dietary Factors

While the general topic of diet also is treated elsewhere (Chapter 14), it may not be amiss to comment on the wisdom of providing for submariners the "best chow in the Navy." There is also evidence that the "open ice box" tradition is a good one, for it appears that a higher and more consistent level of muscular and general efficiency is maintained on five or six meals rather than three meals per day (**57**). Even mild degrees of hypoglycemia may lead to impaired performance, while hyperglycemia may help to offset effects of anoxemia.

Another question concerns nutritional requirements under conditions of chronic stress Most workers are agreed that vitamin increases above the level obtained in a good diet are not helpful under normal conditions (**72**), but there appears some indication that vitamin utilization may be *increased under various types of stress.* Niacin requirements, for example, may increase with elevation of external temperatures (**64, 71**).

III. Measurements of Stress

A. Expressive Manifestations

The fact that in emotion organic changes normally facilitative of adjustment to the environment may occur in the relative absence of overt motor activity suggests that a need for action may be sensed by the individual notwithstanding that he may be aware of the futility of doing anything about it. In other words the emotionally induced organic changes may be symptomatic of effects within the nervous system. For this reason the outwardly observable signs of emotion have often been referred to as "expressive" (**14**). Procedures for recording these peripheral changes have been referred to as "expressive methods" (**35**). Any attempt we make to explore the nature of psychophysiological relationships or to devise tests of psychophysiological performance under stress is likely to employ such expressive methods.

The number and variety of physiological changes which can be recorded by mechanical, optical, pneumatic, electrical and other devices is practically unlimited. In practice what is recorded is determined in part by accessibility of the activity for convenient registration. It may not be assumed, however, that all recordable activities are equally worth registering or that they have the same significance as indices of arousal or emotion. Psychophysiologically the various physiological measures may have quite different significance. Among more frequently used indicators are the following:

1. Electromyograms. The muscles have already been considered at some length as the site of symptoms expressive of psychophysiological "tension" (**28, 68**) and reaction to stress.

2. The Galvanic Skin Reflex. GSR, an index of palmar sweating, is one of the most valuable and readily used indicators. It is one of the most useful indices of "tension" or "background excitation" (**22**), particularly so far as the higher levels of the brain are involved. Functionally, as has been pointed out, palmar sweating is represented in the motor region of the cortex and is accessory to manipulative motor processes in a manner analogous to the way in which salivary secretion may be preparatory and facilitative for the taking of food. It is preparatory for and facilitative of better grip on objects, and is more automatic, dependable, and sanitary than the familiar practice of spitting on the hands. The reaction is neurologically controlled via the sympathetic system from premotor areas of the cortex and mediated peripherally at the sweat glands by acetylcholine. These observations may help account for the fact that the GSR is probably our best expressive indicator for alertness and changes in "attention" (**21**). This measure has provided particularily valuable information bearing on individual differences in stability and the relative excitation level and the rate at which a person will recover following stress (**22, 43, 44**).

3. Blood pressure is another frequently recorded preparatory and facilitative activity (**21**), but it appears less closely related than the GSR to higher level processes of attention etc., and more closely linked with more primitive activities of defense and self preservation. This may explain its value as an index of disturbed emotion. It may account for the fact that it has proved the most useful physiological index in so called "lie detectors" in situations where an individual's life, liberty, reputation and so forth are being defended.

4. Respiration is also easily recorded, but has both the advantage and disadvantage that it is partially under voluntary control. It is primarily an index of the need of the body, real or imagined, for increase of oxygen or elimination of carbon dioxide. The inhibition of respiration or its voluntary regulation following critical emotional stimuli, and the deep sighing inspiration, indicative of respiratory debt when the crucial point has passed, may betray unwary subjects of such tests.

5. Oxygenation of the blood recorded photoelectrically from the ear by an "oximeter" has proved of value in the study of stress, particularly of the kind encountered under conditions of flight (**102**).

6. Vasomotor changes recorded in the extremities or elsewhere by volumetric (plethysmographic), photoelectric (**62**), or thermoelectric methods provide information regarding control of the peripheral vascular bed, metabolic and autonomic (sympathetic) reaction to stimulation.

7. Heart rate, particularly when combined with a record of blood pressure, may provide valuable clues as to shifts in activity of the autonomic system. Sympathetic excitation will tend to increase it, often in association with a rise in blood pressure. A rise in blood pressure, via effects on carotid sinus and aortic pressure receptors, on the other hand, tends to slow it. Accordingly changes in heart rate, taken by themselves, may be misleading (**23**). Sinus arhythmia (**101**) attributable to inhibition of vagal impulses to the heart during inspiration may provide some indication of the amount of the uninhibited parasympathetic tone.

8. Blood sugar, as pointed out by Cannon (**12**) may provide valuable information regarding mobilization of the sympathico-adrenal system. More recently Gellhorn and collaborators (**52**) have demonstrated that the vago-insulin system also participates in response to excitation, its activity being overshadowed by the sympathetic in strong excitation, but its effects being experimentally demonstrable. Its effects often persist following the recession of the sympathetic phase. The latter observation may help explain many exaggerations of parasympathetic activity, effects on the gastro-intestinal system, etc., which tend to follow emotional reactions, which, in their initial phase, obviously involve sympathico-adrenal mechanisms.

9. Metabolism, either basal or relative, may be measured by respiratory methods recording gaseous exchange (**6, 31**). It may be measured in the brain by arteriovenous differences in composition of the blood (**63**). It may be indicated less exactly by one of its peripheral manifestations, the total insensible weight loss attributable to evaporative cooling. The latter method, combined with methods for maintaining balance of the recording scales by continuous automatic compensation (**41**), provides valuable indication of short time changes in metabolic activity under stimulation.

10. The adrenal cortex and related changes has been extensively studied following the work of Long (**77**) and Selye (**93**) and others, in relation to the "alarm reaction" or reactions of stress. Central nervous system influences acting via the pituitary and secretion of adrenocorticotrophic hormone may increase secretion of adrenocortical hormone; the amount of 17-ketosteroids excreted in the urine provides an indication. Ketosteroids in the blood may also cause destruction of lymphocytes, and cell counts may provide an index of this effect. Studies of effects of stress on adrenocortical activity offer considerable promise (**38, 39, 65, 77**).

11. Electroencephalography offers means of determining effects of stress within the brain itself (**26, 27, 76, 98**). Effects of so-called "tension" on the EEG have long been recognized. Generally "tension" has been associated with the presence of low voltage fast activity, most pronounced in anterior areas. Conversely, changes occurring with relaxation often associated with increase of the alpha (10 per second) rhythm are familiar to any electroencephalographer. Changes in sleep with the disappearance of alpha and the occurrence of spindles 12 to 14 per second have also been extensively studied (**59**). Combinations of EEG as an index of cerebral activity with various of the above described indicators of organic change offer a possibility for investigation of cerebral-organic interrelationships under stress (**27, 60, 61**).

12. Other expressive indicators employed in psychophysiological study are described by various authors (**11, 12, 16, 23, 28, 33, 35, 39, 51, 58, 60, 76, 79, 81, 83, 86, 90, 98, 101,** and **102**). Methods of employing these indicators for testing capacity for response and reaction to conditions of stress will be briefly considered.

B. Tests of Reaction to Stress

Psychophysiological tests are desired which will screen out incompetence under stress with an effectiveness comparable with that of test, interview, and trial duty methods now in use. The problem of refining the already screened product of the selection mill sets requirements extremely high. Prospective tests may be classified into four general groups relating to the three types of stress already outlined. Such tests may concern (1) capacity for homeostatic adjustment to conditions such as O_2 lack, (2) capacity for performance under stress of duty, (3) emotional stability and response under stimulation, and (4) tests of chronic status involving conditions of internally maintained activity in the absence of specific stimulation. Tests may combine these conditions of stress singly or in varying proportions. For example we may combine a test of *performance* and of *emotional stability* under conditions of *O_2 deficiency* and observe the persisting effects on *chronic status*.

1. Tests of capacity for homeostatic adjustment are perhaps best illustrated by the various studies of the effects of oxygen lack under conditions of posture, CO_2, temperatures, blood sugar, etc., as cited in various studies (**5, 10, 11, 58, 86, 98,** and **102**). Among the better test indicators may be mentioned critical fusion frequency (cff), dynamic visual field (dvf), visual acuity, word association, handwriting, code substitution, steadiness, and color naming. Failure of response to a signal may be used to indicate loss of consciousness (**102**). These indicators may involve an element of performance which is an outstanding feature of tests of type 2. An outstanding example of a test of homeostatic adjustment is the Dreyer rebreathing test (**30**), widely used in the selection of air pilots in the first world war.

2. Tests of performance under stress of duty, fatigue, emotion, etc. Among the better indicators is the color naming test of Bills (**8**) and various devices of the serial sequence and pursuitmeter type described in the literature. The experiments of Patrick (**89**) involving escape from a variably locked compartment is of this class. Many tests involving frustration and the development of "neurosis" on breakdown of perceptual discrimination belong to this group. (See Section V, 8.)

3. Tests of physiological reaction to emotional stimuli. Such tests may combine the various "expressive" indicators into batteries and attempt to determine the pattern of organic response following various stimuli. Such tests may attempt to determine the sympathetic or parasympathetic and muscular tension components of the individual's emotional response. Studies **19, 22, 72, 75, 81, 94, 100,** and **101** are of this type.

4. Tests of chronic status or conditions of internally maintained level in the absence of stimulation. A number of studies (**59, 90, 100**) are primarily of this type.

Factor analysis of data obtained by batteries of such tests to provide information regarding outstanding factors in human adjustment is mentioned as a promising development. Various authors (**19, 86, 94, 98,** and **101**) have demonstrated possible applications of factor analysis in this field. However evaluated and weighted, the development and perfection of a psychophysiological strain gauge from such components would be a valuable contribution.

IV. The Role of Consciousness in Behavior, Emotion, Affect, and Morale

Emotion is here identified with the bodily changes in excess of, inappropriate to, and too disorganized for effective motor adjustment. "Conscious" emotion is identified with the awareness of those changes. "Affective value" has been inferred for those changes to the extent that they have meaning or value to the individual as the result of previous experience. But this survey would be incomplete if it stopped there and gave no consideration to the part the conscious aspect of these activities may contribute in the reaction to stress or to what in general terms may be designated the *maintenance of "morale."*

If by *morale* we here mean those psychological conditions contributing to effective individual and group behavior on board submarines, and if we assume conscious states and attitudes to be important aspects of those psychological conditions, we may properly inquire concerning their mode of operation. Is consciousness a factor in effective behavior? Is it an influence in breakdown under stress? This obviously is no place for metaphysical discussion or for digression on the dangers of dualistic thinking, but if conscious affect has a real and contributing influence on the submariner's performance under stress, it should be possible to fit it into our thinking on the problem.

A. *A Functional View of Consciousness: Analogy to Radar*

That the sensory awareness of the ongoing bodily processes of the moment provides a background upon which is projected an image or representation of that to which we are then directing attention seems a fair formulation. To develop the analogy further, it appears that consciousness may be likened to the image on a radar or television screen (see Halstead's analogy to an amplifier, **58**), into which is integrated in meaningful poly-dimensional pattern the innumerable sensory contributions by the body's myriads of sensory (including proprioceptive) receptors. By the use of certain controls we may, with practice, shift the image, change the focus, or select the source of stimulation. We may look this way or that, recall this event or that, or concentrate on this detail or that—outside or inside the body or within the range of remembered experience. So far as we are aware, the only deliberate thing we do is somehow to change the composition of this sensory complex. Shifting the eyes, selective tensing of various muscles of postural orientation, or talking to one's self are some of the devices we learn to use for purposes of this control. Even the young child achieves coordination as he masters these devices. By permitting, *pari passu* with the event, observation of bodily changes as they occur, consciousness provides a direct and indirect control of bodily movement within the frame of the shifting environment. It provides this without even the most rudimentary knowledge of the neurophysiological mechanisms which supply the integrated patterns and permit facility in their control. Such a screen is no less essential to voluntary direction of thought and behavior than the cathode ray screen is to the radar operator who, watching the pattern, may manipulate controls and intelligently direct complex electrodynamic functions of which he may have not the slightest comprehension. Without such an observable integrated controllable pattern, purposeful direction of the activity would be intellectually and physically impossible. With such a pattern in what may be thought of as *the screen of consciousness* the submariner may, without knowledge of neurophysiology, intelligently direct his complex sensory, neural, organic, emotional, and affective functions.

More remarkable still, he may, depending on how good a set he has—how good is his "imagery"—reinstate, to varying degrees the images of yesterday; he may review last year's horizons. In so doing he may reinstate not merely a shadowy pattern of yesterday's teleoceptor (visual and auditory) patterns, but the interoceptor (organic) concomitants of those patterns with their emotional and affective implications as well. He may thus not merely compare results of today's with yesterday's scanning of the external world, but he may also evaluate today's view of the external world on the basis of yesterday's favorable or unfavorable organic reactions to it. So important are these reinstatements or repetitions from yesterday that they may, especially with eyes closed, tinge the entire screen of consciousness. Indeed, the submariner's awareness or field of consciousness may be occupied by far different patterns from those projected at him from the inside of a hull under 30 fathoms. On a submarine, within the limited orbit of one's gaze, and amid the deadly repetition of each day's routine, it is true if anywhere, that a man carries his own ultimate resources. In the last resort his escape mechanisms are within himself.

And herein lies the problem of morale. As always on a submarine, a major consideration should be habitability, and this is no less true of psychological habitability than of other kinds. Although each man may carry certain devices for "escape" within his own processes of consciousness, the problem of maintaining morale, stated in general terms, consists in the maintenance of such a high level of psychological habitability in terms of interest, cameraderie, *esprit de corps*, mutual respect, etc., that his limited and permissible leave in the land of fantasy shall be as pleasant as possible. The physiological dividends of such psychological habitability may be assured.

V. Suggestions and Discussion

In the survey of evidence relating to stress questions have arisen regarding possible desirability of changes in policy, procedure and equipment. Some of these possibilities can be immediately disregarded by the experienced submariner as impracticable; some of them may be recognized as appropriate subject matter for future research.

1. Chronic cumulative effects of CO_2. So far as we know the effects of excessive CO_2 over extended periods have not been investigated as have effects of oxygen deficiency. This offers an important unexplored field for investigation. In view of the cerebral vasodilator action of CO_2 and its tendency to favor low voltage fast activity in the electroencephalogram similar to effects of mental effort, "tension," and anxiety, it would be appropriate to study the psychological effects carefully. If there are such effects it would be of interest to know if anticholinergic drugs such as those used for combatting seasickness would counteract the effect without impairment of cerebral circulation or other undesirable effects.

2. Effects of salt and water deficiency at high temperatures. The dramatic and rapid effects of water deprivation at high temperatures on psychophysiological status, as described by Bean and Eichna (**5**) and other workers, suggest it as a means for readily reversible changes in the study of the relation of physical status to readiness, interest, "morale," etc.

3. Temperature control. Needless to say, warm bulkheads decrease the cooling obtained by radiation. At 93 to 95°F. no cooling is obtained by radiation or conduction, and at higher temperatures the body may gain heat through these sources. Insulation or shielding of motors and other heat sources where possible offers obvious benefit. Modern forms of insulation now have eliminated the old fire hazard. Noise might also be thereby reduced. Furthermore, since one 100-watt bulb produces as much heat as one man, it would seem desirable to cut down as far as possible the number of filament lamps and increase the use of fluorescent lighting.

Since the condensate from the air conditioning units is a valuable source of fresh water, it would appear that everyone might benefit by keeping humidity low. There is need for a truly silent air conditioner.

4. Anticholinergic drugs in motion sickness. The use of anticholinergic drugs for the control of motion sickness raises the question whether there may not under certain contions be an increased susceptibility to anoxia and hyperventilation due to effects upon the cerebral blood vessels. Results from EEG studies suggest that possibility (**24, 25**). Research may be indicated.

5. Alleviation of effects of stress. While it is true that chronic and severe states of fatigue and impairment take an inevitable toll of performance and efficiency that must eventually be compensated, it is conceivable that anything that could postpone the day of reckoning would be of value. What can be done for the already dead-tired men aboard when faced with new or continuing emergencies?

Benzedrine, for example, has been widely used to postpone fatigue and drowsiness. Individuals vary widely in any benefits gained in this way, and "parodoxical" effects are frequent. Chemically, benzedrine is rather similar to ephedrine, both stimulating the sympathetic nervous system; blood pressure is increased but blood sugar not affected. Tests of work-to-exhaustion with reflex or more complex activities show that under these conditions more work may be accomplished. Best effects, however, occur in the psychological spheres of boredom, mood, willingness to work, and sociability, with improvement being generally manifested. Little work has been done with chronic fatigue and impairment states using normal subjects, although there are favorable reports dealing with the effect of benzedrine on patients suffering from chronic exhaustion, depression and psychoneurosis.

Caffeine has a direct stimulating effect on muscle (with increased O_2 consumption), reflected in such tests as tapping and hand grip. While steadiness may be decreased (both benzedrine and caffeine make many individuals "nervous"), there may be improved coordination. More prolonged stimulation effects are found on "higher" mental processes such as are involved in color naming, calculation, discrimination, etc. Increased alertness and postponed drowsiness results from ingestion of either benzedrine or caffeine. The deleterious syndrome of "coffee nerves" that plagues the poor civilian is apparently unknown in he Navy: sack duty retains its popularity, even though a war head be used as a pillow.

Smoking in an enclosed area such as a submarine means that all hands, including nonsmokers, are exposed to persistent subacute dosages of nicotine and respiratory irritation. Smoking is known to induce peripheral vasoconstriction, lowering of skin temperature, increase in heart rate and to slightly raise the blood sugar level. Blood pressure effects are variable. There is decreased steadiness (increased tremor) and probably some slight decrease in tests of rote learning and memory span. Because of habituation and addiction such effects vary markedly among individuals; it is probably unwise to attempt to change the well-ingrained smoking habits of a group of men. The strong personal element in fatigue and stress would suggest that if a man *thinks* he gets a "lift" from a cigarette, he should be allowed to smoke—always with due regard to the smoking lamp.

The occasional cautious reference in log books to the effects of a ration of brandy for all hands agrees with informal comments of various submariners; the beneficial and heartening glow is most welcome. Something may be said for it as a sedative for relief of "tension." The eventual depressive effect of alcohol makes it imperative that it be used judiciously, e.g., after the peak of demand for activity has passed and ceremonials are appropriate. It plays a more important role during refits and shore periods.

It would require a more profound (or more presumptuous) treatise than this to offer a "solution" to the age-old conflict between military life and sex life. Thorough screening has eliminated the homosexual, equivalent screening for heterosexual activities would eliminate the Navy; men even love their boats. It is not the purpose of this paper to prescribe for men on shore.

6. Habituation as preparation for stress. There appears to be a sound psychophysiological basis for Naval emphasis on repeated practice drill until virtual perfection is attained in the performance of both individual and group activities. This would tend to reduce energy requirements and any excess tension involved in the activity. Furthermore, some degrees of "overtraining" may result in the maintenance of competent performance despite the presence of a very considerable amount of psychophysical impairment. The point is well illustrated by Halstead's account (**58**) of a concert violinist who, while nearly 'asleep' from excessive fatigue, gave a technically errorless performance; another violinist gave an equally flawless performance under conditions of "acute anoxia." The formal aspects of a highly differentiated skill persisted without a continuously associated high working level of consciousness (**58**, p. 95).

The value of "trial duty" in the testing and training of men is well recognized. To whatever extent practicable, routine undersea experience in the actual encountering of depth bombing and escape from apparently "disabled" craft would seem to be valuable preparation for actual warfare. Psychophysiological study under such conditions might make great contribution to our knowledge of emotion.

7. Drowsiness. There have been several attempts to develop automatic devices based on detectable organic changes for arousing the person who becomes drowsy. It would seem that in a submarine a low-energy radio or wire-connected system by which watchers would respond routinely to an irregularly presented signal from a central station might provide reliable security. Signals might be presented automatically and a lapse of a determined interval without response might automatically give an alarm.

8. Adaptation vs. cumulative effects of excitation. That repetition of excitation may lead to adaptation and indifference or to increasing irritation and even neurosis is a matter of frequent observation. What are the characteristics of the stimuli, aside from intensity, which differentiate them in this respect? We do not know. What are the psychophysiological characteristics of excitation background and response which identify the two types of reaction? We may only guess. Can one type be changed to the other by changing the attitudes or physiological states of the subjects? This is a problem on which there may well be very profitable psychophysiological research.

VI. Summary

An attempt has been made to evaluate evidence related to three types of stress: A, stresses of habitability; B, stresses arising in performance of duty; C, stresses involving emotion following perception of danger; and, in addition, D, conditions of physiological impairment accessory to and aggravating effects of stress. In the course of this presentation special consideration has been given the following:

1. Important differences, attributable to the presence of carbon dioxide, between reactions to low oxygen undersea and at high altitudes.

2. "Side effects" of the normal adjustments to stress which may account for persisting and pathological effects.

3. The roles of background excitation and postural substrate under stress as factors in performance.

4. The view that "emotional" bodily changes differ from normal changes only in that their physiological adjustments exceed requirements and may be excessive, inappropriate, or disorganized.

5. The influence of emotionally induced organic changes upon the brain.

6. The role of consciousness in behavior and morale.

References

1. Adolph, E. F. Physiological fitness for the desert. *Fed. Proc. Amer. Soc. exp. Biol.*, 1943, **2**, 158–164.
2. Arnold, M. B. Physiological differentiation of emotional states. *Psychol. Rev.*, 1945, **52**, 35–48.
3. Bartley, S. H., & Chute, E. *Fatigue and impairment in man.* McGraw-Hill, New York, 1947.
4. Baldock, G. R., & Walter, W. G. A new electronic analyser. *Elect. Eng.*, 1946, **18**, 339–344.
5. Bean, W. B., & Eichna, L. W. Performance in relation to environmental temperature. *Fed. Proc. Amer. Soc. exp. Biol.*, 1943, **2**, 144–158.
6. Benedict, F. F., & Benedict, C. G. Mental effort in relation to gaseous exchange, heart rate, and the mechanics of respiration. *Carnegie Inst. Publ.*, 1933, Publ. 446, pp. 83.
7. Best, C. H., & Taylor, N. B. *The physiological basis of medical practice.* Williams and Wilkins, Baltimore, 1939.
8. Bills, A. G. Fatigue in mental work. *Psychol. Rev.* 1937, **17**, 436–453.
9. Bills, A. G. Facilitation and inhibition in mental work. *Psychol. Bull.*, 1937, **34**, 286–309.
10. Brozek, J. M. Psychological factors in relation to performance and fatigue. *Fed. Proc. Soc. exp. Biol.*, 1943, **2**, 135–144.
11. Burton, A. C., & Taylor, R. M. A study of the adjustment of peripheral vascular tone to the requirements of the regulation of body temperature. *Amer. J. Physiol.*, 1940, **129**, 565–577.
12. Cannon, W. B. *Bodily changes in pain, hunger, fear, and rage.* 2nd Ed. Appleton, New York, 1929.
13. Cannon, W. B. *The wisdom of the body.* W. W. Norton & Co., New York, 1932.
14. Cannon, W. B. The role of emotion in disease. *Ann. int. Med.*, 1936, **9**, 1453–1465.
15. Carr, H. A. *Psychology: a study of mental activity.* Longmans Green & Co., New York, 1925.
16. Case, T. J., Adams, W., Alving, E., Bay, E., Bryan, H., Carmichael, H., Halstead, W., & Landown, M. *Chronic effects of intermittent anoxia in man—Studies of cerebral electrical activity.* OEM, cmr, 113. July 1943.
17. Chorobski, J., & Penfield, W. Cerebral vasodilator nerves and their pathway from the medulla oblongata. *Arch. Neur. Psychiat.*, 1932, **28**, 1257–1289.
18. Courts, F. A. Relations between experimentally induced muscular tension and memorization. *J. exp. Psychol.*, 1939, **25**, 235–256.
19. Darling, R. P. Autonomic action in relation to personality traits of children. *J. abnorm. soc. Psychol.*, 1940, **35**, 246–260.
20. Darrow, C. W. Emotion as relative functional decortication: The role of conflict. *Psychol. Rev.*, 1935, **42**, 566–578.
21. Darrow, C. W. Galvanic skin reflex (sweating) and blood-pressure as preparatory and facilitative functions. *Psychol. Bull.*, 1936, **33**, 73–94.
22. Darrow, C. W. The equation of the galvanic skin reflex curve. I. The dynamics

of reaction in relation to excitation-background. *J. gen. Psychol.*, 1937, **16**, 285–309.

23. Darrow, C. W. Physiological and clinical tests of autonomic function and autonomic balance. *Physiol. Rev.*, 1943, **23**, 1–36.
24. Darrow, C. W., Green, J. R., Davis, E. W., & Garol, H. W. Parasympathetic regulation of high potential in the electroencephalogram. *J. Neurophysiol.*, 1944, **7**, 217–226.
25. Darrow, C. W., & Graf, C. G. Relation of electroencephalogram to photometrically observed vasomotor changes in the brain. *J. Neurophysiol.*, 1945, **8**, 449–462.
26. Darrow, C. W. The electroencephalogram and psychophysiological regulation in the brain. *Amer. J. Psychiat.*, 1946, **102**, 791–798.
27. Darrow, C. W. Psychological and psychophysiological significance of the electroencephalogram. *Psychol. Rev.* 1947, **54**, 157–168.
28. Davis, R. C. Methods of measuring muscular tension. *Psychol. Bull.*, 1942, **39**, 329–346.
29. Davis, H., & Wallace, W. McL. Factors affecting changes produced in electroencephalogram by standardized hyperventilation. *Arch. Neur. Psychiat.*, 1942, **47**, 606–625.
30. Division of Military Aeronautics. *Air Service Medical Manual.* Govt. Printing Office, Washington, 1918.
31. DuBois, E. F. *Basal metabolism in health and disease.* 3rd Ed. Philadelphia, Lea and Febiger, 1936.
32. Duffy, E., & Lacey, O. L. Adaptation in energy mobilization: changes in general level of palmar skin conductance. *J. exp. Psychol.*, 1946, **36**, 437–452.
33. Dunbar, H. F. *Emotions and bodily changes.* New York: Columbia University Press, 1935.
34. Dunn, W. H. Emotional factors in neurocirculatory asthenia. *Psychosom. Med.*, 1942, **4**, 333–354.
35. Dysinger, D. W. A comparative study of affective responses by means of the impressive and expressive methods. *Psychol. Monogr.*, 1931, **41**, 14–30.
36. Eichna, L. W., Ashe, W. F., Bean, W. B., & Shelly, W. B. *The upper limits of environmental heat and humidity tolerated by acclimitized normal young men working in hot environments.* Proj. 2-11, 2-13, 2-17. File No. 727-2, Fort Knox, Ky. Open.
37. Finesinger, J. E., *et al. Studies on the electroencephalographic changes induced by anoxia.* Office of Naval Research, Contract N5 ori-76, Proj. XIV 05-664, Sect. I, 5-20.
38. Finesinger, J. E., *et al. Endocrine studies,* Office of Naval Research, Proj. Order XIV 05-664, Contract N5 ori-76, Sect. IV, 50-61.
39. Finesinger, J. E., & Brazier, M. A. B. *Studies on the physiological basis of the EEG changes found in hypoxia and their comparison with changes induced by other agents.* Office Naval Research, Proj. order XIV 05-664, Contract N5 ori-76, pp. 1–16.
40. Freeman, G. L. The facilitative and inhibitory effects of muscular tension upon performance. *Amer. J. Psychol.*, 1933, **45**, 17–52.
41. Freeman, G. L., & Darrow, C. W. Insensible perspiration and the galvanic skin reflex. *Amer. J. Physiol.*, 1935, **111**, 55–63.
42. Freeman, G. L. The postural substrate. *Psychol. Rev.*, 1938, **45**, 324–334.
43. Freeman, G. L. Toward a psychiatric plimsoll mark: Physiological recovery quotients in experimentally induced frustration. *J. Psychol.*, 1939, **8**, 247–252.
44. Freeman, G. L., & Pathman, J. H. The relation of overt muscular discharge to physiological recovery from experimentally induced displacement. *J. exp. Psychol.*, 1942, **30**, 161–174.
45. Fulton, J. F. *Physiology of the nervous system.* New York: Oxford University Press, 1943, 2nd Ed.
46. Gellhorn, E., & Spiesman, I. G. The influence of hyperpnea and of variations on O_2- and CO_2- tension in the inspired air upon hearing. *Amer. J. Physiol.*, 1935, **112**, 519–528.
47. Gellhorn, E., & Kraines, S. H. Word associations as affected by deficient oxygen, excess of carbon dioxide and hyperpnea. *Arch. Neur. Psychiat.*, 1937, **38**, 491–504.
48. Gellhorn, E. The influence of carbon dioxide in combating the effects of oxygen deficiency on psychic processes with remarks on the fundamental relationship between psychic and physiologic reaction. *Amer. J. Psychiat.*, 1937, **93**, 1413–1434.
49. Gellhorn, E. Oxygen deficiency, carbon dioxide and temperature regulation. *Amer. J. Physiol.*, 1937, **120**, 190–194.
50. Gellhorn, E. Circulatory studies on anoxemia in man with respect to posture and carbon dioxide. *Ann. int. Med.*, 1937, **10**, 1267–1278.
51. Gellhorn, E., & Kessler, M. The effect of hypoglycemia on the electroencephalogram at varying degrees of oxygenation of the blood. *Amer. J. Physiol.*, 1942, **136**, 1–6.

52. GELLHORN, E. *Autonomic regulations—their significance for psychology and neuropsychiatry.* New York: Interscience Publishers, 1943.

53. GIBBS, F. A., GIBBS, E. L., & LENNOX, W. G. The value of carbon dioxide in counteracting the effects of low oxygen. *J. Aviat. Med.*, 1943, **14**, 250–261.

54. GERARD, R. W. Brain metabolism and circulation. *Proc. Res. nerv. & ment. Dis*, 1938, **18**, 316–345.

55. GRINKER, R. R., & SPIEGEL, J. P. *Men under stress.* York, Pa.: Maple Press, 1945.

56. HAGGARD, E. A. Some factors determining adjustment during and readjustment following experimentally induced stress. In: *Contemporary Psychopathology.* S. TOMKINS, ed. Harvard U. Press, 1943. See: *J. exp. Psychol.*, 1943, **33**, 257–284.

57. HAGGARD, H. W., & GREENBERG, L. A. *Diet and physical efficiency.* New Haven: Yale Univ. Press, 1935.

58. HALSTEAD, W. C. *Brain and intelligence. A quantitative study of the frontal lobes.* Chicago: Univ. Chicago Press, 1947.

59. HENRY, C. E. Electroencephalographic study of individual differences and their constancy: I. During sleep. *J. exp. Psychol.*, 1941, **29**, 117–132.

60. HENRY, C. E., & DARROW, C. W. Autonomic factors in the relation of EEG to heart rate. *Fed. Proc. exp. Biol.*, 1947, **6**, 128–129. A basis for interpreting autonomic EEG relationships. *Ibid.*, 1947, **6**, 93–94.

61. HENRY, C. E., & DARROW, C. W. EEG relationships to five autonomic variables. *Amer. Psychologist*, 1947, **2**, 421.

62. HERTZMAN, A. B. The blood supply of various skin areas as estimated by the photoelectric plethysmograph. *Amer. J. Physiol.*, 1938, **124**, 328–340.

63. HIMWICH, H. E., BOWMAN, K. M., DALY, C., FAZEKAS, J. F., WORTIS, J., & GOLDFARB, W. Cerebral blood flow and brain metabolism during insulin hypoglycemia. *Amer. J. Physiol.*, 1941, **132**, 640–647.

64. HOWELL, W. H. *Textbook of physiology.* Fulton, J. F., editor. Phila.: Saunders, 1946.

65. HOAGLAND, H. Changes in the EEG and in the excretion of 17 Ketosteroids accompanying electro-shock therapy of agitated depression. *Psychosom. Med.*, 1946, **8**, 246–251.

66. HUNT, W. A. Recent developments in the field of emotion. *Psychol. Bull.*, 1941, **39**, 249–276.

67. IVY, A. C. High altitude problems in aviation. *Fed. Proc. Amer. Soc. exp. Biol.*, 1946, **5**, 319–327.

68. JACOBSON, E. *Progressive relaxation.* (2nd ed.) Chicago: Univ. Chicago Press, 1938.

69. KENNARD, M. A. Autonomic interrelations with the somatic nervous system. *Psychosom. Med.*, 1947, **9**, 29–36.

70. KEYS, A., MATTHEWS, B. H. C., FORBES, W. H., MCFARLAND, R. A. Individual variations in ability to acclimatize to high altitude. *Proc. Roy. Soc. Lon.*, 1938, **126**, 1–29.

71. KEYS, A. *Diet and fatigue in hot climates.* Medical nutrition report #13, 1943.

72. KEYS, A. Physical performance in relation to diet. *Fed. Proc. Amer. Soc. exp. Biol.*, 1943, **2**, 164–187.

73. KLEITMAN, N. *Sleep and wakefulness.* Chicago: Univ. Chicago Press, 1939.

74. KOTTKE, F. J., PHALEN, J. S., TAYLOR, C. B., VISSCHER, M. B., & EVANS, G. T. Hypoxia and temperature regulation of mice, dogs and man. *Amer. J. Physiol.*, 1948, **153**, 10–15.

75. LASSWELL, H. D. Certain prognostic changes during trial (psychoanalytic) interviews. *Psychoanal. Rev.*, 1936, **23**, 241–247.

76. LINDSLEY, D. B. Electroencephalography. In: *Personality and the behavior disorders.* Edited by J. McV. Hunt. New York: Ronald Press Co., 1944.

77. LONG, C. N. H. The conditions associated with the secretion of the adrenal cortex. *Fed. Proc. Amer. Soc. exp. Biol.*, 1947, **6**, 461–484.

78. LUND, H. *Emotions; their psychological, physiological and educative implications.* New York: Ronald Press, 1939.

79. LURIA, A. R. *The nature of human conflicts.* New York, Liveright, 1932.

80. MAGNUS, R., & KLEIJN, A. DE. Die Abhängigheit des Tonus der Extremitätenmuskeln von der Kopfstellung. *Pflüg. Arch. ges. Physiol.*, 1912, **145**, 455–548.

81. MALMO, R. B., SHAGASS, C., DAVIS, J. F., CLEGHORN, R. A., GRAHAM, B. F., & GOODMAN, A. J. Standardized pain stimulation as controlled stress in physiological studies of psychoneurosis. *Science*, 1948, **108**, 509–511.

82. MARGOLYN, S., KUBIE, L., KANZER, M., & STONE, L. Acute emotional disturbances in torpedoed seamen of the Merchant Marine who are continuing at sea. *War Med.*, 1943, **3**, 393–408.

83. MASSERMAN, J. *Principles of dynamic psychiatry.* Phila.: Saunders, 1948.

84. McCulloch, W. S. *Machines that think and want.* Detroit Symposium on Psychophysiology and Abnormal Psychology. W. E. Halstead, Ed. (In press)

85. McFarland, R. A. *The effects of oxygen deprivation (high altitude) on the human organism.* Report #13, Dept. Commerce, Bureau of Air Commerce, Safety and Planning Division. 1938, pp. 1–89.

86. McFarland, R. A., Graybiel, A., Liljencrantz, E., & Tuttle, A. D. An analysis of the physiological and psychological characteristics of two hundred civil airline pilots. *J. Aviat. Med.*, 1939, **10**, 2–52.

87. Murphy, J. P., & Gellhorn, E. The influence of hypothalamic stimulation on cortically induced movements and on other action potentials of the cortex. *J. Neurophysiol.*, 1945, **8**, 341–365.

88. Murphy, J. P., & Gellhorn, E. Further investigation on diencephalic-cortical relation and their significance for the problem of emotion. *J. Neurophysiol.*, 1945, **8**, 431–447.

89. Patrick, J. R. Studies in rational behavior and emotional excitement. II. The effect of emotional excitement on rational behavior in human subjects. I. Rational behavior in human subjects. *J. comp. Psychol.*, 1934, **18**, 153–195.

90. Peterson, W. F., & Levinson, S. A. *The skin reactions, blood chemistry and physical status of "normal" men and of clinical patients.* Chicago, 1930.

91. Rosenblueth, A., Wiener, N., & Bigelow, J. Behavior purpose and teleology. *Phil. of Sci.*, 1943, **10**, 18–24.

92. Schmidt, C. F. Intrinsic regulation of the cerebral circulation. *Arch. Neurol. Psychiat.*, 1937, **37**, 1233–1235.

93. Selye, H. *Textbook of endocrinology.* Montreal: Univ. Montreal, 1948.

94. Sherman, M., & Jost, H. Frustration reaction in normal and neurotic persons. *J. Psychol.*, 1942, **13**, 3–19.

95. Sharp, L. H. Effects of residual tension on output and energy expenditure in muscular work. *J. exp. Psychol.*, 1941, **29**, 1–22.

96. Shock, N. W., & Hastings, A. B. Studies of the acid-base balance of the blood. IV. Characterization and interpretation of displacement of the acid-base balance. *J. biol. Chem.*, 1935, **112**, 239–262.

97. Serota, H. M., & Gerard, R. W. Localized thermal changes in the cat's brain. *J. Neurophysiol.*, 1938, **1**, 113–124.

98. Thorner, M. W. *A study of cerebral physiology at high altitudes.* Randolph Field, Rep. 271, open 6144.

99. Winslow, C. E. A. Man's heat exchange with his thermal environment. In *Temperature, its measurement and control in science and industry.* New York: Reinhold, 1941.

100. Wenger, M. A., & Irwin, Orvis C. Fluctuations in skin resistance of infants and adults and their relation to muscular processes. *Univ. Iowa Stud.*, 1936, **12**, 143–206.

101. Wenger, M. A. Studies of autonomic balance in Army Air Forces personnel. *Comp. Psychol. Monogr.*, 1948, **19**, 1–110.

102. Wood, E. H., Lambert, E. H., Baldes, E. J., & Code, C. F. Effects of acceleration in relation to aviation. *Fed. Proc. Amer. Soc. exp. Biol.*, 1946, **5**, 327–351.

Chapter 21

PSYCHOLOGICAL CAUSES AND RESULTS OF STRESS

ERNEST A. HAGGARD

University of Chicago

Introduction

Emotional stress and high morale, while radically different in some respects, are nevertheless similar in others. Emotional stress is experienced by an individual when his general well-being is jeopardized in an emergency situation which threatens the satisfaction of his basic physiological or psychological needs. In such a case, the individual's capacity to keep from going to pieces is severely taxed and he tends to collapse. He is, or feels he is, unable and disinclined to cope with the situation. The experience of fear and anxiety, the disruption of various physiological and psychological functions, and the disintegration of habitual skills and techniques for dealing with reality are characteristic reactions in a stress-producing situation. Breakdown will result if the stress is sufficiently intense and enduring. Morale, on the other hand, reflects an optimal state of adjustment in which the individual experiences a strong sense of belonging to the group, has zest and enthusiasm for his work, and feels a secure confidence in his ability to master reality, both present and future, and desires to do so.

In considering the various aspects of problems involved in the causes and results of emotional stress, the individual will be thought of as an integrated, self-regulating system of abilities and needs, with consequent strivings for their satisfaction (cf. **14, 53**). This concept applies to various levels of complexity, ranging from the basic bodily needs to such acquired psychological needs as self-esteem and acceptance by one's fellows. On each level, the individual is so built that he tends to maintain a relatively stable state of equilibrium or adjustment.

Any severe or prolonged interference with the satisfaction of needs on any level will ultimately threaten the integration of the total individual. When this happens, the individual characteristically gives "danger signals"—such as hunger, pain, anxiety, or neurotic symptoms—which indicate that all is not well. Basic stress-producing threat may result from situational factors, from disturbances in the person's inner adjustment, or from difficulties in the individual's relation to a situation and his inability to master it.

Causes of Emotional Stress

In terms of a given situation, there are many fundamental or predisposing causes within the person which may underlie the development of emotional stress. Apart from obvious physical, mental or emotional deficiencies, a wide range of individual differences may play a real, even though hidden, role in the person's ability to tolerate stress. Basic energy reserves and thresholds for exhaustion, undue sensitivity to intense stimuli, the effects of differences in glandular functions on endurance, etc., are some of the factors which might throw the balance one way or another in a critical situation.

Personality Structure

One important but little investigated factor is the possible effect of an individual's personality structure (or character structure) on the development of emotional stress in a given situation. Personality structure determines, to a large extent, the subtle and complex manner in which the person perceives both the inner and the outer world (e.g., in a cold, intellectual manner, or with deep and stirring feelings), how he reacts to it (e.g., by being outgoing or withdrawn, by accepting or rejecting it), what aspects of it have meaning and value for him, and are of primary importance (e.g., people, things, ideas), the nature of his goals, and the activities which give him basic satisfactions, or a feeling of being lost, misplaced and futile.

Numerous approaches have been used to appraise and describe personality or character structure. Some of the ways of looking at it include the analysis of personality traits and of the things which one values (**1**), physical and temperamental types (**85**), categories of dominant needs and emotional demands (**64**), and the canalization of energies into stable response patterns (**63**). As for its origin, character structure may be due in part to inherited or native tendencies, but it is certain that it is dependent to a large extent on what things happen to the individual from early childhood on, and how he reacts to them.

The Development of Identification Patterns

As the child develops into a socialized individual, a conflict exists between the child and his parents. The child wishes to obey his "instinctual" impulses, whereas the parents wish to impose restraints. However, if the child feels sufficiently loved and secure he will come to "internalize," i.e., impose upon himself the wishes and regulations of the parents. In this process, the child yields (or learns to control) part of his own wishes in order to retain the love, approval and acceptance of the parents. In time, he chooses to do what he has been required to do. That is, he "identifies" with the parents and their value systems.

By thus conforming to authority, the child both retains the love and acceptance of the parents, on which he is dependent for his physical and emotional well-being, and at the same time preserves his "ego" and self-respect. The child must also make similar adjustments to brothers, sisters, and playmates. As a result of internalizing all these social demands and restrictions, the child's character, conscience, ideals, and value systems are moulded. This process typically continues up through adolescence to maturity, so that the individual's character is influenced by all those upon whom he depends for security and approval. As an adult, the individual tends to react to other authority figures (e.g., the boss, family, society, God, etc.) in a similar manner and for similar reasons and rewards.

The socialization process does not always proceed smoothly and without mishap. If as a child the individual feels basically rejected, and finds that he receives nothing for giving up his impulses and wishes, he will tend not to become socialized (e.g., be able to postpone gratifications and assume responsibility for his behavior) or develop a stable character structure (e.g., possess functioning values and goals which are socially acceptable); rather, he will tend to reject society and be unwilling to conform to current authorities. Such persons are generally called psychopaths. Or, again, if, as a child, the person resented and hated his father, or was strongly ambivalent to him (i.e., both loving and hating him), these feelings will tend to be re-enacted in his adult relationships with his superiors. In the same manner, reactions to brothers, sisters, and playmates are reflected in later life. Thus, the adult's character structure is strongly influenced by the type of persons with whom he identified as a child, and the nature of his reactions to these identification figures (**32**).

Identification Patterns, Adult Behavior, and Emotional Stress

Individuals may reveal peculiarities which drastically interfere with their efficiency and adjustment when placed in a stress-producing situation. Their difficulties do not lie in any lack of skills and abilities, but rather are due to temperamental or personality factors. Why is it that some can take commands without resentment, can cooperate with others and strive for the benefit of the group, whereas others cannot? A large part of the answer lies in the relation between the individual's character structure, his early identification patterns, and the various demands and gratifications provided by the situation.

One means of minimizing emotional stress is to evaluate carefully the nature of the situation, and to place individuals with compatible character structures in it. If the person is properly placed, so that he lives out his ego and group identifications, life is meaningful and gratifying (**28**), but if not, constant latent tensions permeate his emotional life and behavior. In the latter case, he may not know why he is not happy or satisfied in his work, but only that he feels at odds with himself and others, that something is wrong. Sufficient and protracted conflict between the potentialities of character structure and the demands of reality in itself may result in neurotic breakdown (**76**) and delay the individual's recovery (**78**).

Emotional Stress Aboard Submarines

A necessary aspect of military life in general, and submarine life in particular, is the relation between the officers and the enlisted men. The officer gives orders, is dominant, and looks after the welfare of the men; the enlisted men take orders and are dependent on the officers for their well-being and safety.

Both officers and men must be able to accept their respective roles if they are to work together with maximal efficiency. In order to avoid emotional stress and breakdown, the officer must be emotionally able to assume and exhibit responsibility (**58**), be thoroughly and competently trained (**73**), show fairness and consideration in dealing with his men (**34**) and feel genuine concern for their welfare without getting too emotional (**58**), and should treat and respect them as individuals. In other words, to the men he should represent a strong, competent, yet benevolent father figure (**57**), and should be able to maintain his role on the basis of his own character strength, rather than solely on the basis of external status symbols, such as rank. Correspondingly, the men must have faith and confidence in the ability, training, and maturity of their officers (**73**), must feel that their officers are concerned about their (the men's) welfare when the situation calls for it (e.g., placing a badly injured man in officers' quarters (**7**)), and must be able emotionally to accept the authority and directives of the officer, and identify with him and his goals (**21, 34, 71**).[1]

Difficulties arise, not only when the officer is unable to carry out his parental role, but also when the enlisted man is unable to shift earlier parental (and family) ties in order to establish the proper and necessary officer (and group) identifications. This inability may be due to several reasons, such as overly strong dependencies on early family attachments (**29**), the experience of too much rejection or pampering as a child (**66**),

[1] From a series of interviews with submariners who had several war patrols to their credit, it appeared that these men wanted two general qualities in their officers. In the first place, they wanted officers who would watch out for their safety, well-being, and comfort. This meant, among other things, that the officers should know their business thoroughly, know their boat 'like a book,' and know how to handle her under all possible conditions. Furthermore, the officers should see to it that the men were properly taken care of in all ways—whether it involve a good menu, a 4.0 rating for the boat, or seeing to it that the men's health was taken care of—and who would not, for example, run unnecessary risks (with their lives) just to get an award. The second general quality had to do with the officers' strictness, consistency, and fairness in dealing with the men, in seeing to it that everyone did his work properly, and that all equipment was always in perfect working condition. In short, these men seemed to say, "If you will protect and take care of me, I will do whatever you ask."

or because the individual has actually grown too mature (or independent) to find satisfaction in returning to a dependent childlike role (**81**), or is unable to do so without anxiety.

Similarly, the group relations among the men should be considered. Under conditions ideal for the group, the men must break some of their emotional ties to their family and home, and shift them to their group. In a sense, then, this new group takes the place of the family. As (emotional) compensation for this shift of ties, and for yielding some of their independence, they feel a sense of belongingness with the group, of esprit de corps (**57**). After shifting their identifications they are able to submerge some of their own selfish interests for the welfare of the group (**34**), and at the same time they are provided with a realistic orientation to the present situation. They then find it possible to share their possessions with their "buddies," or even risk personal injury for them, as they once would for members of their family. The inability to sever family ties, to submit to new forms of discipline, to change one's way of life—that is, the inability to readjust emotionally is an important factor in causing breakdown (**19**). Consequently, the men should be made to feel a member of the group, especially if it is small (**21, 72**), and, if necessary, techniques should be used to create this feeling (**11, 13, 71**). If, on the other hand, the roots of the individual's identifications with the officers and the group are not deep and emotionally satisfying, the frustrations and deprivations, the submission to authority and regimentation, the fatigue, the stress of battle, the fear of death, etc., will undermine his motivations to fight and survive. As a result, he will tend to become resentful and antagonistic and to seek some form of escape.

As a rule, the greater the emotional stress experienced by all members of the group, the more they become welded together, and the more the welfare of the group supersedes the individual's. This is perhaps why, when one member cracks under stress, the composure of the others is severely threatened—almost as if their own defenses and integrative capacities were shattered. Thus, whereas group identifications generally serve to forestall breakdown under stress (**42**), other members of the group may be more vulnerable (**97**) when one of them does go to pieces (**73, 98**).

Over and above the ability to feel oneself a member of the fighting group, the individual's ability to form positive identifications in a context which is broader than that of the family-like group is an additional factor which may influence his susceptibility to breakdown under emotional stress. Examples include ties and allegiance to one's country, or to political, social or religious values. It was frequently observed that individuals possessing such values tended to be much less apt to suffer emotional breakdown as a result of war experiences (**8, 10, 16, 30, 56, 68, 82, 92, 95**). The individual's values and beliefs should be consistent with the action required of him (**4, 6, 72**). This more abstract type of identification is essentially based upon, and is a generalization of, the more concrete forms discussed above. Perhaps identification with these value systems stabilizes the individual by serving as a buffer to present stresses, by making it worth while to overcome temporary hardships for a goal that is more important than his own personal feelings and concerns.

Besides observations of psychiatrists during war-time, various findings have been reported which bear on the relationship between early identifications and adult adjustment. In one study, concerned with analyzing subjects' reactions to several types of leader-follower groups (**48**), it was observed that the ability to accept and like a strong, dominant leader is apparently dependent on the individual's having experienced an early home atmosphere which was both firm and warm. Another investigation was directly concerned with relations between family background and adult work adjustment (**32**), in which the latter was defined in terms of the individual's over-all adjustment, the amount of satisfaction that he was able to find in his work, and his general capacity to use his job qualifications. Some of the significant differences between a well- and poorly-adjusted group are as follows: as children, the former experienced little feeling of rejection as a child, developed strong positive affections (rather than antagonisms) for their parents, identified with a stable, mature adult, did not suffer from excessive sibling rivalry, and in general felt little ambivalence toward their early home situation, which was characterized by an atmosphere of family loyalty and unity. The reverse pattern obtained for the poorly-adjusted adults. Such non-emotional factors as mere economic security made no appreciable difference.

The early attitudes and adjustment patterns of both groups tended to continue into their adult home and work relationships. For example, the well-adjusted individuals were more able to tolerate temporary delays and frustrations and assume responsibility for their actions and lives, showed a stronger tendency to struggle in the face of difficulty, were less reliant on "pull" to get ahead, showed less tendency to have accidents or to spoil their job chances, were more able to use their abilities and skills in their work, and in general showed a more objective, realistic attitude in the selection of and preparation for their work. Again, the reverse tendencies obtained for the poorly-adjusted individuals.

Reactions to Stress and Impulse Mastery

As a general rule, severe emotional deprivations and frustrations during childhood, or the previous experience of traumatic (emotional) shock in similar situations, is an important factor in breakdown due to emotional stress (**29, 66, 90**). This statement is based on the finding that individuals who previously had neurotic symptoms will tend to develop more crippling disabilities when subjected to severe stress, that breakdown (as in combat) is essentially a reactivation of early vulnerabilities by some appropriate stimulus situation (**2, 44, 81**). Various studies of breakdown incidence of service men have reported that from 50 to 97 percent (usually about 80 per cent) of the patients had suffered previous personality disturbances (cf. **5, 33, 55, 57, 70, 77, 88, 98**).[2]

But man is not entirely at the mercy of traumatic shock experiences or his neurotic symptoms. In fact, it was not an uncommon finding that individuals who had traumatic childhood experience, or came from broken homes, or had histories of neurotic difficulties, did not break down under the stress of battle, but on the contrary turned in excellent performances (**12, 30, 55**). The staff of the OSS assessment program (**68**) also found a striking frequency of such trauma and complexes in the life histories of very effective people. In many cases it appeared that such complexes were related more directly to the proficiencies than to the deficiencies of the individual's personality.

The nature of the individual's reactions to both external stress and the degree of his control over emotional impulses may hold part of the answer to this question. As mentioned earlier, a stress or emergency situation is one in which the individual's welfare, safety, and integrative capacities are threatened, and in such cases, the individual mobilizes energies to cope with the threat. If properly directed, these energies enable him to escape or to fight. On the psychological level, the individual usually feels fear (or anxiety), aggression, or both. Fear, the warning that threat exists, may be alleviated by its removal; aggressive impulses

[2] Exclusion of a person on the basis of the 'stop questions' used in psychiatric interviews (cf. **60**) and in the various personal inventories (cf. **97**) seems justified insofar as the symptoms indicate the individual's inability to resolve or control some underlying emotional difficulty which might interfere with his efficient performance of duty.

provide the energy and incentive to achieve this.

It furthermore seems a fundamental characteristic of the functioning of any living organism that such mobilized energies, irrespective of their specific nature, seek expression. From the standpoint of one's inner economy, the most efficient ways of reacting to the stress-producing situation is either to escape or to attack the source of threat. But since in wartime it often happens that neither is possible, it is necessary that the person have adequate means of controlling such impulses as fear and aggression.

If, in the past, the individual was punished and forced to hold in check the expression of aggressive impulses, so that stable patterns of self-assertion were not developed, he will tend to be anxious and guilt-ridden when aggression is called for (**42**). If his inadequate control of such impulses collapses in the face of the stress, he may in consequence be overwhelmed with anxiety—he may feel guilty because of his aggressions—and may fall back into earlier and more infantile modes of behavior. Or, if aggressive impulses do not find adequate and satisfactory outlets, they tend to express themselves less adaptively, as in developing chronic anxiety, a sense of effort or fatigue, physical illness, or in unconsciously causing accidents, or in other forms of neurotic behavior.

Thus, it is necessary that the individual who is to be placed in threatening situations be able to adjust to and tolerate emotional stress. One consequence of the failure to do so is the presence of accumulated and mounting emotional tensions, a condition which indicates that the person's controls and adaptive mechanisms are not functioning properly. If this state of affairs is allowed to continue, breakdown will result, since there is a limit to how much any man can withstand (**20, 22, 23, 79, 83**).

Evidence for the desirability of expressing aroused energies is seen in the results of two experimental frustration studies. In one (**31**), it was found that the individuals who work off their aroused energies in some overt activity, even though it is apparently of no use (nonadaptive), also recover inner (autonomic) equilibrium more rapidly after the stress is terminated. This finding was confirmed in a subsequent study (**37**); and it was also found that subjects who directed their activities toward the relevant problem-task at hand during the period of stress recovered inner equilibrium more rapidly than those who engaged in random, irrelevant actions. Similarly, in numerous animal studies, it was found that a too rigid blocking of activity in stress situations facilitated the development of "experimental neuroses" (cf. **3, 47, 51, 52, 84**).[3] These findings illustrate the every-day observation that it is both difficult and enervating to do nothing when one is excited to activity. Familiar examples of this include aviators for whom it was more difficult to "sweat out" a mission than to go on one, or submariners who, while being depth charged, were forced simply to submit to it.

The release or expression of pent-up energies alleviates the inner tensions they cause. This is perhaps why submariners, after a depth charging episode, usually delight in surface battle after it is over, and then feel "satisfied" about things. Thus, when men under stress cannot express their aroused energies directly, they should be provided

[3] The justifications for utilizing animal studies in this report is twofold. First, the physiological complexity and functional adequacy of the autonomic and vegetative mechanisms, which are involved in emotional-adjustive behavior, are generally as satisfactory in lower animals as in man, so that generalizations of such findings to human behavior seems justifiable. (This is not true to the same extent of behavior involving the central nervous system and the higher mental processes.) Second, the use of animals in experimentation permits a degree of knowledge of causal relationships resulting from the control of the subjects' past and present experiences which is not possible with human subjects.

with some form of substitute activity as an outlet or release. If possible, such substitutes should be symbolically appropriate to the inhibited activities (as in bowling where one aggresses against the pins), or be useful in preserving security and safety (as a daily work period on a submarine). An experimental analysis of substitute aggressive behavior (displacement) in rats (**59**) examines many of the conditions under which such behavior occurs and forms which it might take.

Development of Impulse Mastery in Adults

Although the groundwork for the ability to master such impulses as fear and anxiety is laid in childhood, partial control may be developed in adults lacking it by gradual exposure to stress-producing situations. That is to say, it is possible for adults to become "seasoned" or "toughened" to emotional stress so that it may be tolerated without undue strain. If, for example, an individual is required to perform his duties in a moderately stressful situation, and if over a period of time the stress is gradually increased, he should become able to master the increased stress without excessive difficulty. As a consequence, and within reasonable limits, there should be no appreciable loss in his ability to function adequately under conditions that previously would have been incapacitating. By thus becoming progressively adapted to the stress situation, the individual learns that by keeping his head he will be subjected to minimal danger and at the same time will be able to act most effectively. Realistic training under realistic conditions of stress is a necessary procedure if many individuals are not to become psychiatric casualties when confronted by actual danger. In fact, one observer believes that "shell shock" is nothing but insufficient training (**73**).

A related procedure involves the forcing of a successful solution under conditions of conflict or stress. This method is illustrated by an experiment with cats (**52**) which were made neurotic by induced conflict between hunger and fear of the feeding situation. When the animals were in a state bordering on panic, they were forced by their intense hunger to break through their previous fear of the food box. Once this occurred, signs of neurotic behavior subsided, and the previously traumatizing stimuli came to be a signal for feeding. This "sink-or-swim" method, however, is not without its possible risks, such as when the stress is too great, or when the subject is unable to achieve a satisfactory solution. After this experience of forced solution, some of the animals entered into a state of panic, and continued to show a marked increase in their neurotic reactions, both in and out of the experimental setup.

The proper mental and emotional attitudes of the individual also may aid in the mastery of stress-producing situations. In an experiment involving conditioned responses to strong electric shock (**35, 36**), the subjects who actively tried to face and master the situation (as opposed to those who remained passive) showed much less physiological and emotional disturbance throughout the experiment and especially in anticipation of the shock, were less bothered by the shock itself, and were better able to withstand it as the experiment progressed. These subjects likewise recovered their composure more rapidly after the shocks were terminated.

The importance of a person's actively and realistically trying to master the situation confronting him also applies to longer periods of time. For example, in a work adjustment study (**32**), the dominant tendency of the better adjusted individuals was actively to face their difficulties and problems (rather than to feel self-disparagement and the desire to quit), to be more realistic and flexible in their attitudes toward themselves and their work, and to be more willing to risk present disappointment in striving toward their long-range goals.

The Collapse of Mastery and Breakdown from Stress

If the individual's ability to tolerate emotional stress, and his mastery of undesirable impulses are inadequate and collapse, the breakdown may occur either relatively early, or after an extended period of time, or after the stress-producing situation is past (**24, 97**). Early breakdown is apparently due to such predisposing factors as a basic conflict between the person's identification patterns or character structure and the demands of the situation (**67**), or to inadequate control of powerful and disturbing emotions. When an individual breaks down after a long period of successful adjustment, it appears that the precipitating causes are primarily inherent in the stressful characteristics of the situation (**9**). When breakdown occurs after the stress situation has terminated, it appears due to a relaxing of the individual's defenses against such impulses, and a return of the anxiety previously held in check, which then overwhelms him.

Reduction in the number of breakdowns of the first type can largely be achieved by proper selection and placement procedures (**89, 93, 94**). The second and third types of breakdown can be avoided in many cases by watching for the patterns of symptoms which typically precede such "situation neuroses" (**41, 65, 81**). If the symptoms of an individual suffering from this latter type of stress are recognized in time, and he is temporarily relieved to allow for early treatment (**43, 53**) and recuperation, or is reassigned to other duties, the frequency of breakdown as a result of prolonged or recurrent emotional stress should be reduced sharply.

Cognitive Structure and Emotional Stress

Generally speaking, the individual's cognitive structure of a situation refers to his mental picture of it. The more accurate and complete this picture is, the better the cognitive structure. This factor is relevant to emotional stress, since a person is able to act realistically and effectively in such a situation only if he knows the nature and seriousness of the threat, knows what to do, and is able to do it.

The presence or absence of cognitive structure partially determines the difference between fear and anxiety. If a person knows what to expect in a threatening situation, he tends to experience fear—that is, fear of something which is more or less specific, localized, and which can be dealt with. On the other hand, anxiety, the dread of the threatening unknown, is diffuse and unlocalized, hence is generally incapacitating, since an individual full of uncertainty is unable to act realistically and effectively (**18, 27, 38, 39, 41**).

A necessary condition of good cognitive structure is that the individual possess adequate information covering all aspects of the situation and his relation to it. In a war, this would include knowledge of the threat potential of the enemy, the adequacy of the defenses against him (both detection and protection), the predictability and effectiveness of one's own aggressive strength, and the extent of the advantage and risk involved in each of several possible paths of action (**4, 38, 54**).

Emotional stress is intensified when cognitive structure is lacking, for example, when an individual is uncertain of the effectiveness of the enemy's means of detecting and attacking him (e.g., if he does not know whether the position of his boat has been located, or if the expected depth charge will come, and if it will be a direct hit). Stress is similarly increased if he does not know the adequacy of his own means of detection of and protection against the enemy (e.g., whether he will be surprised by an enemy dive bomber, or if the boat will strike a floating mine, or if it is strong enough to survive the depth charging, or can be taken down fast enough to escape or avoid being seen) as well as the effectiveness of his aggressive measures (e.g., whether the torpedo will find its target, or perhaps even circle back upon the boat). The average submariner's need for cognitive structure is great, because in effect he has yielded all possibility of developing it himself, since those functions are consigned to the officers (**28**). Furthermore, anyone engaged in combat always fears that the dreaded unexpected will happen, even after long experience (**39, 69**).

During the recent war, there were opportunities to appraise the effect of cognitive structure on emotional stress. When enlisted men were told the purpose and risk involved in a particular operation, there was both an improvement in performance and a corresponding decrease in breakdowns and general friction among the men (**7, 16, 17**).[4] This was also found to be true in industry during the war. In one case, a general orientation course about the plant was used to show the importance of the man's job in relation to the total war effort, and brought about a rise in morale and work output and a drop in absenteeism.

An additional aspect of cognitive structure is concerned with the individual's available skills and techniques for acting effectively in a situation. For maximal protection against emotional stress, the individual should have experienced all the possible contingencies that might arise, and should have learned how to handle them. It is not enough just to have read or heard about what to do in a crisis—the person should have experienced a less extreme version of the situation in training, and learned what to do. This ideal may be impossible to achieve in practice, but in any case the individual should be so instructed during training that he will not be at a complete loss as to what to do, and have to fall back on his imagination. To have been through such experiences, and to have developed responses which are quickly and automatically executed makes the individual both more self-confident and more efficient in his work—at a time when seconds may be of vital importance (**18, 24, 27, 38**).

A sense of helplessness results, on the other hand, when an individual is caught off guard, or untrained to act effectively in the situation (**24, 39, 41**). Panic frequently follows. This inability to act as a result of lack of cognitive structure is illustrated by the aviator's anxiety over the unpredictable and uncontrollable, such as flak, and was dreaded much more than enemy fighters, which could be seen and shot at. Another example is the submariner's anxiety when having to wait in suspense for depth charges without being able to escape or retaliate.

Several experimental studies are relevant to the problem of the effect of lack of cognitive structure on the sense of helplessness. In one (**52**), cats were trained to master a conflict situation by learning how to control the feeding signals. These subjects overcame their neurotic inhibitions and other symptomatic behaviors in the situation. With such mastery, normal, adaptive behavior was resumed under the same physical conditions which had previously produced neurotic behavior. Similarly, in a related study (**62**), two groups of rats were given the same amount of physical punishment (electric shock), with only one group having mastery over its duration. By several indices, the various behavior disturbances manifest by the helpless subjects did not diminish as the experiment progressed from day to day, whereas the others apparently

[4] In interviews, submariners reported that on a war patrol, they had a strong need to know as much as possible about everything that was going on. For example, if a sister boat had been sunk nearby, they wanted to know about it, so as to be on guard all the more; or if one had had successes, they wanted to rejoice with her. The unnecessary withholding of information by officers made the men feel isolated, anxious, and in a sense rejected, and they assumed that the officers did not think their interests or feelings worth considering. This necessarily hindered their identification with the officers as interested, benevolent father figures, and the resulting anxieties tended to interfere with their morale and efficient performance.

adapted to it, so that they ceased to be bothered as they were at first. In an experiment with human subjects (**35, 36**), one group had mastery over a strong electric shock which was used to produce fear and anxiety, in the sense that they both knew when it was coming and controlled its administration. The other group lacked such cognitive structure and control. All other aspects of the situation were the same for both groups. In this study, the former subjects showed less inner physiological (autonomic) arousal, were less disturbed by the shocks, were less anxious in anticipation of the shocks, and adapted to them more easily. In addition, various overt responses were disrupted less for these subjects, and in retrospect they showed less cognitive distortion in their description of various technical aspects of the situation.

Motivational Patterns and Emotional Stress

Generally speaking, motivation initiates, directs, and sustains an individual's behavior, and is a critical factor in his adjustment to emotional stress (**7**). If a person is required to perform a certain task, it is important to know both whether his motivations for doing it are strong and enduring. It is also important to know the quality or nature of his motivational patterns. For example, one may be motivated to carry out his assigned task, but underneath may also be driven to do the opposite (e.g., be aggressive against the authority figure, or punish himself), so that he is torn in conflict, ambivalent, and hence inefficient and undependable in the stress-producing situation.

A series of motivations found to go hand in hand with emotional stability and dependability for hazardous duty (**68**) include the following: sociocentric motives, such as the desire to contribute to the welfare of the group or society, or to help achieve a commonly shared goal, and motivations based on realistic confidence in one's ability and training to do the work involved. Undesirable motivations, found in unstable individuals, are egocentric and stem from the desire to escape from something distasteful (as from work, home or personal problems), for superficial status motives (financial prestige), or from attempts to gain self-respect through short-cuts (e.g., to ease a guilty conscience).

Emotional stress is likely to result if the individual's motivations are inappropriate, inadequate, or in conflict, in terms of the demands and gratifications of a particular situation. In such cases, the individual does not feel satisfied in his work; he feels frustrated and unhappy, and is basically unwilling and unable to do his part in the face of stress.

Inappropriate motives are often characterized by being too narrow or specific to cover the range of activities that must be engaged in. In submarine warfare, if the men are motivated only to sink enemy shipping, they will have a feeling of success when this is accomplished; if they fail to do so, they will feel disgruntled, demoralized, and personally inadequate. However, under the same objective conditions, the feeling of success could be achieved, either by broadening and making more flexible the dominant motivational patterns, or changing the whole outlook, or adding substitute goals (cf. **45**). Substitute behavior is more effective when similar goals (rather than similar acts) are involved (**15**). Futhermore, if one has identified with other individuals or groups, he may gain basic satisfactions when they achieve the desired goal (**46**). Thus, if during training or briefing, men were shown motion pictures of an effective air-sea rescue, or really became convinced of the importance of reconnaissance work, and if good rapport existed between the various branches of the service, the sense of drudgery and futility would not weigh so heavily on men engaged in this type of duty. Rather, such work would become both important and satisfying.

Monotonous, repetitious, routine work which at the same time requires close attention is frequently marked by inadequate motivations. One example is seen in men who operate anti-submarine sound gear, whose motivation to carry out the particular activity over a long period of time seems to become expended. If this condition (usually called satiation) continues long enough, the person will show strain, shifts of attention, fatigue, and lowered efficiency. If it continues under high tension, emotional breakdown and personality disintegration may result (**8, 99**). Satiation develops most rapidly when the individual feels he is repetitiously marking time and not accomplishing anything, and it is dispelled by either temporarily shifting to other activities, or by feeling that real progress is being made (cf. **45**).

Emotional stress may result from a conflict of motives. Such conflict may lie at deep as well as at superficial levels. For instance, a person may select a line of work for what he thinks is a particular set of motives; but this conscious assumption may only cover up deeper contradictory motives, such as a desire to humiliate one's parents, prove one's manhood, aggress against authority figures, etc. (**32**). In such cases, the individual who is prevented from satisfying these deeper emotional needs will tend to feel frustrated, torn apart, and unable to do his work effectively. Other examples of hidden conflicts were mentioned earlier, such as the need to express aggression, along with feelings of guilt and anxiety over it. In many cases, the person may be aware of the conflict, as when he does wish to do his duty, but at the same time is driven to avoid risk of injury, disability or death. Such emotional conflicts are inevitable in war, but if they are not resolved or mastered, they will play havoc with the person's adjustment and performance.

An individual's motivation and ability to perform effectively in a stress situation may be interfered with as a result of debilitating physiological factors. These generally contribute toward a depletion of energy reserves, a disintegration of adaptive capacities, and hence increase susceptibility to breakdown. Examples include excessive fatigue, insufficient sleep, extreme variations in temperature and humidity, poor and insufficient nutrition, infections, injuries, illness, toxic conditions, etc. (**24, 91**). These factors often operate subclinically—i.e., they are not clinically evident—so that their influence may not be realized even by the man himself. However, this does not minimize their importance as a contributing factor in lowered efficiency or even breakdown as a result of emotional stress (**21, 24**).

Situation Conditions and Emotional Stress

Besides the various factors within the individual which have been mentioned, there are several situational conditions which may precipitate the collapse of the individual's adjustive mechanisms. In general, the analysis of such situational factors has been made on the basis of investigations with animal subjects (cf. footnote 3), and is discussed in the literature on "experimental neuroses." Reports and summaries of conditions under which such behavior disorders may be experimentally induced through environmental stress may be found in various sources (cf. **3, 47, 51, 52, 84**).

There are several common factors which contribute toward the development of non-adaptive behavior in animal subjects. They are: excessive and overwhelming stimulation, being caught off-guard after a series of less intense stimuli, or specific susceptibilities of certain types of subjects. A similar effect results from the presentation of stimuli which recur too frequently or for too long a time to permit adequate adjustment or rest; the monotonous repetition of a specific pattern of stimulation with trivial, irrelevant consequences. This condition generally exists when there is a strict adherence to a highly rigid time schedule, even though its duration is not excessive; the protracted inhibition or forced delay of a response after the signal to act is given. The suspense and strain of such delay may be either self-imposed or the result of environmental restraints; forcing the subject to alternate between action and restraint in response to a particular stimulus pattern; forcing the subject to make an impossible discrimination between "positive" and "negative" stimuli (i.e., those followed by "reward" or "punishment"); or a sustained conflict between two antagonistic drives. In such cases, the subject must be strongly motivated, and at the same time prevented from evading the insoluble problem; similarly, on the action side, forcing the subject to respond when no specific response is available may result in breakdown.

In concluding this section, it will be noted that several of these factors correspond to the various conditions underlying the development of emotional stress discussed earlier in this chapter, such as inadequate cognitive structure or motivational conflict. The additional point should be emphasized that all of the conditions which underlie susceptibility to breakdown of an individual's adjustive mechanisms as a result of emotional stress are mutually interdependent. For example, if an individual is emotionally insecure and anxious, he will find it difficult to evaluate the situation clearly and to determine the most effective path of action. On the other hand, if he does not know what to do in an emergency situation, he will tend to become anxious and insecure. Consequently, with regard to a potentially stress-producing situation, considered attention should be paid to each of the various factors, in order to minimize emotional stress and maximize efficient performance.

Results of Emotional Stress[5]

Emotional stress, if sufficiently intense and prolonged, results in the disruption of one or more of the adaptive functions of the organism. Such dysfunction occurs when the intricate mechanisms of adjustment are taxed beyond their normal limits of tolerance. The various symptoms of disturbance which appear under such conditions indicate that the individual's resources for mastering the stress have been surpassed. The duration of the dysfunction may be either temporary or permanent, depending on such factors as the nature and severity of the stress, the particular vulnerabilities of the individual, the adequacy of his integrative mechanisms and means of handling stress, and whether he gains any advantage from his symptoms. The focal point of the disability may involve the physiological, behavioral, mental, or interpersonal levels of adjustment, or various combinations of these.

Physiological symptoms resulting from emotional stress are called psychosomatic. This type of disturbance may take the form of a marked increase or decrease in the activity of the affected organ system, and may or may not involve irreversible structural changes. To understand and deal with psychosomatic symptoms, both the physiological changes and their emotional correlates must be considered, since in many cases they go hand in hand.

The normal functioning of any self-regulating system which serves to maintain the

[5] Relatively little space will be given in this chapter to a consideration of the results of emotional stress, for three reasons. First, the available information dealing with the results of breakdown from stress is both more extensive and more adequately formulated. For obvious reasons, students of human behavior cannot systematically study the various factors underlying the collapse of an individual's adjustive mechanisms by imposing sufficient stress. Rather, they must make their observations after the breakdown has occurred. Second, even though one understands the results of emotional stress, it is usually not possible to reconstruct completely the individual's futile attempts at adjustment, or retrace the complex pattern of events and reactions that led up to the collapse of his adaptive behavior. That is to say, we cannot learn from the results of breakdown very much about the causes. Third, in terms of dealing with a practical situation, the important thing is to understand and be able to control the factors in the individual and in the situation that might produce emotional stress, and hence to prevent breakdown.

physiological equilibrium of the organism may suffer from excessive emotional stress. The literature in this field is extensive and well-documented (cf. **25, 53, 60, 80, 96**), and there is substantial evidence that more or less specific emotional syndromes are frequently associated with disruptions of the following physiological systems: circulatory (essential hypertension, cardiac neurosis, Raynaud's syndrome, neurocirculatory asthenia, effort syndrome); respiratory (bronchial asthma, vasomotor rhinitis); gastrointestinal (anorexia nervosa, nervous vomiting, bulimia, visceroptosis, nervous diarrhea, chronic psychogenic constipation, duodenal and peptic ulcers, spastic colitis, mucous colitis, ulcerative colitis); genitourinary (nocturnal enuresis, frequency of urination; difficulty in voiding, impotence); endocrine (hyperthyroidism, hyper- and hypoglycemia, exophthalmic goiter, obesity); locomotor (neuritis, lumbago, fibrositis); the skin (eczema, pruritus, psoriasis, urticaria, rosacea complex); the central nervous system (headache, migraine, chorea, insomnia, weakness and faintness, epilepsy); and various special senses (eye: asthenopia, glaucoma, miners' nystagmus, conjunctivitis; ear: vertigo, and hypersensitivity to all forms of stimuli.

Similarly, emotional tensions resulting from stress may effect a wide range of overt behavior disturbances. The causes underlying the various symptoms, as well as their nature, extent, and duration, have a common denominator, regardless of the level of adjustment on which they appear. In each case, they result from inadequate mastery and utilization of mobilized energies to attack or escape from some form of threat to the well-being of the organism. The apparent difference between the psychosomatic and the behavioral symptoms is that the energies or impulses are inhibited in the former, whereas they are expressed overtly in the latter (**26, 87**).

Under the impact of sufficient stress, the customary integrated, flexible, adaptive responses of the individual tend to become fragmented, inflexible, and nonadaptive in terms of the requirements of the reality situation. Although the possible number of such overt behavior disturbances is legion, they tend to fall into patterns, such as: the loss of precise, integrated, coordinated responses (e.g., tremor, tenseness); disorganized behavior (e.g., panic); irrelevant activities (e.g., excessive smoking); inappropriate reactions (e.g., phobic responses in an objectively safe situation); misdirected expression of aroused impulses (e.g., displaced aggression); impulsive or compulsive behavior (e.g., accident proneness, tics); inhibition or repression of overt expression (e.g., catatonia, hysterical paralysis); stereotyped, constricted, rigid behavior, which is generally non-adaptive (e.g., fixations); and a return to previously adequate and adaptive behavior patterns (e.g., regression, primitivation of action). Such means of responding in a stress-producing situation are to some extent adaptive. They provide the individual with at least some form of expression of his aroused energies, and hence partially relieve inner tensions.

Certain findings from studies of the phenomena of fixation and regression may be utilized to advantage in preparing men to act effectively in a potentially stress-producing situation. Fixated behavior, which is highly stable, rigid, and resistant to modification, characteristically occurs under conditions of strong motivation or emotional stress. The strength of fixation is also a function of how often the particular act has been performed, and the immediacy and degree of resultant satisfaction gained by the individual (cf. **84**). That is to say, response patterns previously employed in a stress-producing situation recur persistently on subsequent occasions. Fixations are highly adaptive if the required performance of the individual remains unchanged, but are equally nonadaptive if he must modify his behavior under conditions of stress, since he will very likely do again just what he did on the previous occasions.

Behavioral regression occurs when the individual, faced with a highly distasteful or seemingly intolerable situation, finds that his response patterns do not remove the threat, and reverts to some previously learned mode of reacting to such situations. However, behavioral regression does not occur under such conditions unless the individual has previously learned to do something different in the situation (**61**). Thus, it is important that men who must act automatically and appropriately in the face of stress should be well trained under the very conditions in which stress is likely to develop, in order to establish "fixated responses." Furthermore, the required tasks should be stable and not be inconsistent over a period of time, in order to avoid regression under stress to previously adaptive, but now non-adaptive, responses.

Under conditions of stress, the higher mental processes may be so distorted that they cease to serve the individual adaptively. There is ample evidence in the clinical and experimental literature (cf. **63, 74**) to show that the various cognitive processes involved in a person's relation to himself (as seen in such defense mechanisms as rationalization, projection, self-deception, the repression of undesirable wishes, thoughts, and experiences) may be drastically affected as a result of unresolved motivations, emotional tensions, and anxieties. For similar reasons, the mental processes involved in dealing with the external environment (e.g., attention, perception, thinking, learning, remembering) also may be drastically distorted by such conscious and unconscious forces. The individual's mechanisms of adjustment suffer in both respects if the emotional stress is sufficiently intense and prolonged.

An individual's adjustment in terms of his interpersonal relationships may be non-adaptive as a result of emotional stress. The weakening or collapse of the integrative mechanisms may result from past or present events, or may be a consequence of situational factors or be due to disturbances within the individual. Typical maladaptive reactions in relation to others may involve excessive and unrealistic tendencies toward withdrawal or isolation from others, dependence on others, or aggression against other individuals (cf. **40**). Similar disturbances of interpersonal relationships (as well as the other levels of adjustment) have been observed in laboratory animals subjected to stress sufficient to produce maladaptive or "neurotic" behavior patterns (cf. **3, 47, 51, 52, 60**).

In conclusion, in appraising the causes and results of emotional stress, it is important to keep in mind that the total individual is involved in his relation to the stress-producing situation, and in his attempted adjustment to it. This is particularly true of the various disturbances of overt behavior, the higher mental processes, and the interpersonal relationships. In these cases, the connection is rather loose between the cause of the breakdown of the individual's adaptive mechanism and the various behavioral manifestations, and in addition, if one area is involved, it is likely that the others are also. The psychosomatic disturbances, however, may show a fairly close tie between the pattern of the individual's emotional adjustment and the manifest symptoms, which in turn may not involve, to any appreciable degree, disturbances on the more complex levels of adjustment. But in any case, the individual afflicted by the breakdown of any of his adaptive mechanisms is less in command of his native and acquired abilities to cope with and master the stress-producing situation, so that his over-all adjustment is impaired.

Some Research Proposals

Analysis of the Submarine Situation

In the first place, it seems highly desirable to make a complete, explicit, and objective analysis of all aspects of the submarine situation. This analysis should include the activities and conditions that characterize life aboard a submarine, such as the realistic (physical) situation into which the men are placed and to which they must adjust, the particular spacial and temporal demands made upon the submariner, the kind of emotional and physical tolerances and skills needed, the dynamics of the groups that exist aboard the submarine, the degree of responsibility and level of intellectual and emotional maturity required for each activity, and the special rewards and satisfactions that the men receive. Such an analysis should not be limited to the submarine situation of the present, but also should be oriented in terms of future conditions, such as the effect of longer periods submerged and other changes resulting from mechanical improvements.

Selection and Placement to Avoid Emotional Stress

If a completely adequate description of the submarine situation were available, it should be possible to deduce the essential characteristics of men who would adjust as submariners with minimal emotional stress. That is to say, men should be selected for submarine duty whose objective qualifications, potentialities for adaptation, conscious and unconscious needs, motivational patterns, and identifications are optimally satisfied by the demands and satisfactions provided by life aboard a submarine. To take an oversimplified example: if the situation required that a man possess an IQ of about 120, be about 20 years old, have strong latent aggressive tendencies, a desire for close, protected quarters, and a strong need for group support, then men filling these qualifications, and who possess the necessary skills, should be most likely to adjust satisfactorily as submariners. However, a person with a need to express his aggressive tendencies by independent action, and who felt anxious in close quarters, probably would be dissatisfied and ineffective as a submariner, even though he met all the other qualifications. He might, however, become an excelient fighter pilot.

In an attempt to gain some insight into the character structure and dominant underlying motivation patterns of submariners, a modified Thematic Apperception Test was administered to a group of men who had been on several war patrols, who liked their life aboard submarines, and who had adjusted successfully to it.[6] From a preliminary analysis of the stories to 10 pictures, some of the characteristic latent tendencies that appeared included: a need for change in the environment in order to be happy; an acceptance of the *status quo*, with no strong need to strive for and achieve great things; a superficial need for independence, with a deeper dependence on outside support; frequent unresolved conflict, ambivalence, or fluctuation of aggressive and sexual drives; and a reliance on external stimulation to determine their thoughts, which tend to be specific and concrete rather than elaborate and abstract. It is possible that the universality of these motivational patterns will not be substantiated by future research with submariners; but even so, it is not difficult to see how these tendencies go hand in hand with the requirements of life aboard a submarine.

It is often implicitly (and mistakenly) assumed that there is a universal or ideal pattern of adjustment, and that deviations from it are indicative of some form of maladjustment. A more nearly correct and practicable approach is to be concerned with the particular requirements and gratifications of a given situation, the corresponding motivational patterns of individuals who must adjust to it, and the proper "fit" of the two. Furthermore, if specific motivational patterns can be established as being generally common to successful submariners, it is important that they be appraised in non-value terms, and not as being either "good" or

[6] Approximately 200 stories were recorded by the author at the Mare Island Naval Base.

"bad" patterns of adjustment in any moralistic sense.

The purpose and goal of selection and placement is to assign men to tasks which maximally satisfy their dominant (conscious and unconscious) needs and motivations, and at the same time do not activate difficulties in their adjustment by irritating particular character weaknesses.[7] A person may adjust successfully to some situations, even though he is a misfit in others. For example, an accident-prone individual might be placed to advantage in a commando or paratroop unit, where he could satisfy his tendency to find excitement in risk—but not aboard a submarine. Again, an individual whose only point of vulnerability is a dread of being alone should not necessarily be disqualified for submarine duty, because the occasion which would arouse his anxiety would not exist in this situation. On the contrary, his need for group support may even facilitate his adjustment as a submariner.

Some Suggested Areas for Further Investigation.

From a consideration of the various causes of emotional stress discussed earlier in this chapter, it is apparent that in many ways our present knowledge of these factors is inadequate. Before the conditions underlying the development of emotional stress can be understood adequately and controlled, a number of intensive and extensive investigations of these conditions must be carried out. More specifically, some of the problems which need further study are concerned with:

1. *The nature of the basic personality factors which contribute toward an individual's becoming a successfully adjusted and effective submariner.* This problem involves several aspects, such as whether for such men common factors were present in: (a) their early home and family life, their identification patterns, and their characteristic attitudes and modes of reacting to emotional stress, (b) their ability to sever early emotional dependencies on their family and to transfer them to a contemporary adult group composed of officers and men, or to identify with some broad ideological frame of reference, and (c) their healthy development of emotional stability and "snafu" tolerance in the face of emergency situations.

2. *The stability of the motivational pattern (or character structure) as a result of exposure to the physical and emotional demands of a given situation.* Wheareas it may be that men who have been in submarines for a substantial period of time show characteristic motivational patterns, the question may be raised as to whether such consistency was present when these men volunteered for submarine duty. It may well be that constant exposure to (and necessity of adjustment to) the situation forced the divergent individual patterns to converge upon the generally prevalent pattern. If this be the case, it is important to discover the common factors of adaptability in the character structure of the men who were able to shift their motivational patterns to conform to the requirements of the submarine situation, as well as the possible factors which were present in those who proved unable to make this shift. Two related problems are concerned with (a) whether some degree of initial divergence from the common pattern is desirable, and (b) the nature and extent of the initial divergence that may be tolerated so that the necessary readjustments may be made without undue emotional stress.

3. *The nature of the adjustive mechanisms or integrative capacities which enable an individual to tolerate and master emotional stress.*[8]

[7] An analysis of some of the major problems involved in the selection and placement of personnel, and some suggested solutions to them, have been presented in descriptions of the OSS assessment program (**50, 68**).

[8] The following example illustrates the effectiveness of adaptive mechanisms for handling severe emotional stress: "One man on an antiaircraft gun shook so at the first airplane raid

It was mentioned earlier that about 80 percent of the individuals who suffered breakdown from the stress of battle had previous records of personality disturbances. However, many individuals whose case histories indicated breakdown performed well in military service, whereas about 20 percent of the psychiatric casualties, who were earmarked by psychiatrists and various "personal inventories" as adjusted and stable individuals, collapsed under comparable conditions of stress. With such predictive discrepancies, it is important to discover the nature of the positive resources of the former (regardless of the past traumatic experiences or present neurotic symptoms), and the liabilities and deficiencies in the personalities of the latter individuals. A full answer to this important problem apparently must wait until investigators pay as much attention to the factors that contribute toward resilient mental health and emotional stability as they have to the causation of the neuroses and other personality disorders.

4. *The most efficient techniques for training adults to master emotional stress on the one hand, and to develop maximally effective skills for dealing with the emergency situation on the other.* The point was made earlier that in training men to control themselves and act effectively in a crisis, exposure to simulated emergency situations is often an effective way to "toughen" men to stress. However, it is necessary to determine, for example, (a) the range of individual differences among the men in their ability to tolerate stress, (b) the limits of the stress to be used in training (e.g., from where it is meaningless to where it causes breakdown), and (c) the optimal rate at which stress is to be increased during training, in order to facilitate adaptation to it, and still avoid unnecessary breakdown. Other areas worth investigating include the development of appropriate substitute activities ("safety valves") to avoid the building up on excessive tensions from stress, and various techniques to facilitate effective performance in order to avoid a feeling of helplessness.

5. *The most efficient techniques for providing adults with appropriate motivational patterns and a well-developed cognitive structure of the situation.* The possible effectiveness of various education procedures to develop a broad, flexible, cohesive set of motives should be determined. The nature of the submariner's motivations should not be ignored or left to chance, since to a large extent his morale and performance on the job depend on whether he has a feeling of success and accomplishment. Such positive feelings occur when his motivations are satisfied. It is probably the case that, ordinarily, other members of the group, various extraneous events, etc., determine the individual's motivations to a large extent—but if this condition is left uncontrolled, it may lead to unfortunate results. In some instances, it may be necessary to modify the individual's motivations (e.g., from egocentric to sociocentric, or to make them less rigid and unrealistic), as a contributing factor to the efficiency of the men and the group.

Similarly, the best means of providing the men with an adequate mental picture of the total situation and their place in it should be determined and utilized. Such a program would both enable the men to deal more effectively with the realistic emergency situation, and at the same time be less susceptible to breakdown from emotional stress.

6. *An analysis of factors which tend to increase or decrease mastery of emotional stress in a crisis.* A number of group and situational factors may influence the development of emotional stress in submarines. Systematic investigations should be made of the various factors which might play a critical role in affecting the morale, efficiency, and performance of the men in a crisis situation. Some of these factors apparently include the

that he fell off his seat. He wept, vomited and lost control of his bladder and bowels. But he was the only man who could operate this radar-controlled gun. And he stuck it out through twelve more raids" (**81**, p. 275).

effect of (a) placing men with compatible or incompatible character structures, interests, etc., in a group (**100**), (b) the roles played by the officers and men, whether aloof and disciplinary, as in the German submarines, or the more warm and flexible relations generally found in American crews, and (c) the methods used in dealing with a crisis situation, as for example whether the desirable effects of silent running during a depth charge attack are more than offset by the undesirable effects of forced inactivity, heat, and humidity.

7. *The possibility of training officers in the techniques of handling the group-dynamics among the men.* This general problem includes the officers' skill in recognizing and utilizing the men's ego-, group-, and officer-identifications for the advantage of all members of the group, of instilling feelings in the men that they are individuals worthy of respect, and that their contribution to the total effort is meaningful and important. The emotional forces involved here may determine whether the men surmount, or succumb to, severe emotional stress in a crisis. If properly organized and directed, these energies provide additional strength to master the stress; if disorganized and undirected, they may render the individual virtually defenseless and helpless in the face of severe stress.

Summary

An individual experiences emotional stress when his over-all adjustment is threatened, when his adaptive mechanisms are severely taxed and tend to collapse. Some of the factors which influence an individual's ability to tolerate and master stress include: the nature of his early identifications and his present character structure, and their relation to the demands and gratifications of the present stress-producing situation; the nature of his reactions to the situation; his ability to master strong and disturbing emotional tensions; the extent to which he knows about all aspects of the situation, so that he is not helplessly unaware of the nature and source of threat; his available skills and other means of dealing effectively with it; and the strength and pattern of his motivations to do so.

Under the impact of sufficient stress, response patterns which are customarily integrated, flexible, and adjustive tend to become nonadaptive in terms of the needs of the individual and the demands and gratifications of the external reality situation. That is to say, excessive stress results in the disruption of one or more of the self-regulating, adaptive functions of the organism, and may involve the mechanisms of physiological equilibrium, various forms of overt behavior, the higher mental processes, or the individual's interpersonal relationships. The breakdown of such adaptive functions further impairs the over-all adjustment of the individual.

References

1. Allport, G. W. *Personality: a psychological interpretation.* New York: Henry Holt, 1937.
2. Anderson, C., & Jeffery, M. Psychiatric casualties from the Normany beach-head. *Lancet*, 1944, **247**, 218–221.
3. Anderson, O. D., & Liddell, H. S. Observations on experimental neurosis in sheep. *Arch. Neurol. Psychiat.*, 1935, **34**, 330–354.
4. Anon. Group of 25 Scientists. *Science in war.* New York: Penguin Books, 1940.
5. Ballard, S. I., & Miller, H. G. Neuropsychiatry at a Royal Air Force center: analysis of 2000 cases. *Brit. Med. J.*, 1944, **2**, 40–43.
6. Bartlett, F. C. *Psychology and the soldier.* Cambridge, Eng.: Cambridge Univ. Press, 1927.
7. Bassan, M. E. Some factors found valuable in maintaining morale on a small ship. *Bull. Menninger Clin.*, 1947, **11**, 33–42.
8. Bettelheim, B. Individual and mass behavior in extreme situations. *J. abnorm. soc. Psychol.*, 1943, **38**, 417–452.
9. Bleckwenn, W. J. Neuroses in the combat zone. *Ann. intern. Med.*, 1945, **23**, 177–183.
10. Bortz, E. L. Adjustment in wartime. *Ann. intern. Med.*, 1943, **19**, 457–469.

11. Bradford, L. P., & Lippitt, R. Building a democratic work group. *Personnel*, 1945, **22,** 2–12.
12. Brown, J. A. C. Frustration and aggression. (A study of frustration in relation to Army life.) *J. R. Army med. Cps.*, 1945, **85,** 101–104.
13. Burns, J. H. The psychologist looks at the army. *Infantry J.*, 1928, **33,** 592–602.
14. Cannon, W. B. *The wisdom of the body.* New York: Norton, 1932.
15. Child, I. L., & Grosslight, J. H. The effect of substitute activity as depending upon the nature of the similarity between substitute and original activity. *Amer. J. Psychol.*, 1947, **60,** 226–239.
16. Cohen, R. R. Mental hygiene for the trainee. *Amer. J. Psychiat.*, 1943, **100,** 62–71.
17. Cohen, R. R., & Delanos, J. G. Subacute emotional disturbances induced in combat. *War Med.*, Chicago, 1945, **7,** 284–296.
18. Connors, E. M. The psychological effects of infantry weapons. *Infantry J.*, 1930, **36,** 486–489.
19. Crighton-Miller, H. General conclusion. (Chap. XI) In: Miller, E. (Ed.), *The neuroses of war.* New York: Macmillan, 1940, pp. 201–210.
20. Culpin, M. A weekend with the war neuroses. *Lancet*, 1940, **239,** 257–259.
21. Daly, C. D. A psychological analysis of military morale. *Army Quart.*, 1936, **32,** 59–75.
22. Debenham, G., Sargent, W., Hill, D., & Slater, E. Treatment of war neuroses. *Lancet*, 1941, **240,** 107–109.
23. Dillon, F. Neuroses among combatant troops in the Great War. *Brit. med. J.*, 1939, **2,** 63–66.
24. Duff, I. F., & Schilling, C. W. Psychiatric casualties in submarine warfare. *Amer. J. Psychiat.*, 1947, **103,** 607–613.
25. Dunbar, H. Flanders. *Emotions and bodily changes: a survey of literature on psychosomatic interrelationships.* (2nd Ed.) New York: Columbia Univ. Press, 1938.
26. Dunbar, Flanders. Medical aspects of accidents and mistakes in the industrial army and in the armed forces. *War Med.*, Chicago, 1943, **4,** 161–175.
27. Eltinge, LeR. *Psychology of war.* Fort Leavenworth: Press of the Army Service Schools, 1915.
28. Erikson, E. H. Ego development and historical change. *The psychoanal. stud. of the Child*, New York: International Universities Press, 1946, **2,** 259–396.
29. Fairbairn, W. R. D. The war neuroses; their nature and significance. *Brit. med. J.*, 1943, **1,** 183–186.
30. Farrell, M. J., & Appel, J. W. Current trends in military neuropsychiatry. *Army med. Bull.*, Wash., 1944, **78,** 44–50.
31. Freeman, G. L., & Pathman, J. H. The relation of overt muscular discharge to physiological recovery from experimentally induced displacement. *J. exp. Psychol.*, 1942, **30,** 161–174.
32. Friend, J. G., & Haggard, E. A. Work adjustment in relation to family background. *Appl. Psychol. Monogr.*, Stanford Univ. Press, 1948, No. 16.
33. Gowan, L. R. Psychiatric aspects of military disabilities. *Nav. med. Bull.*, Wash., 1943, **41,** 129–137.
34. Grinker, R. R., & Spiegel, J. P. *Men under stress.* Philadelphia: Blakiston, 1945.
35. Haggard, E. A. Some conditions determining adjustment during and readjustment following experimentally induced stress. In: S. S. Tomkins (Ed.), *Contemporary psychopathology: a source book.* Cambridge: Harvard Univ. Press, 1943.
36. Haggard, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomic reactions during and following experimentally induced stress. *J. exp. Psychol.*, 1943, **33,** 257–284.
37. Haggard, E. A., & Freeman, G. L. Reactions of children to experimentally induced frustration. *Psychol. Bull.*, 1941, **38,** 581. (Abstract)
38. Hall, G. S. Morale in war and after. *Psychol. Bull.*, 1918, **15,** 361–426.
39. Hastings, D. W., Wright, D. G., & Glueck, B. C. *Psychiatric experiences of the Eighth Air Force.* New York: J. Macy Jr. Foundation, 1944.
40. Horney, Karen. *Our inner conflicts: a constructive theory of neurosis.* New York: Norton, 1945.
41. Kahn, L. A. A discussion of some causes of operational fatigue in the Army Air Forces. *Psychol. Bull.*, 1947, **44,** 34–53.
42. Kardiner, A. *The traumatic neuroses of war.* New York: Hoeber, 1941.
43. Kardiner, A., & Spiegel, H. *War stress and neurotic illness.* New York: Hoeber, 1947.
44. Kaufman, M. R., & Beaton, L. E. A psychiatric treatment program in combat. *Bull. Menninger Clin.*, 1947, **11,** 1–14.
45. Lewin, K. Behavior and development as a function of the total situation. In:

L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley & Sons, 1946, pp. 791–844.

46. Lewis, H. B. An experimental study of the role of the ego in work: I. The role of the ego in cooperative work. *J. exp. Psychol.*, 1944, **34**, 113–126.
47. Liddell, H. S. Conditioned reflex method and experimental neurosis. In: J. McV. Hunt (Ed.), *Personality and the behavior disorders*. (Vol. I.) New York: Ronald Press, 1944, pp. 389–412.
48. Lippitt, R., & White, R. K. The "social climate" of children's groups. In: R. G. Barker, J. S. Kounin, & H. F. Wright, *Child behavior and development*. New York: McGraw-Hill, 1943, pp. 485–508.
49. MacKinnon, D. W. A topological analysis of anxiety. *Character & Pers.*, 1944, **12**, 163–176.
50. MacKinnon, D. W. Some problems of assessment. *Trans. N. Y. Acad. Sci.*, Series II, 1947, **9**, 171–185.
51. Maier, N. R. F. Animal neuroses. In: P. L. Harriman (Ed.), *Encyclopedia of psychology*. New York: Philosophical Library, 1946, pp. 33–38.
52. Masserman, J. H. *Behavior and neurosis*. Chicago: Univ. Chicago Press, 1943.
53. Masserman, J. H. *Principles of dynamic psychiatry*. Philadelphia: Saunders, 1946.
54. Maxwell, W. N. *A psychological retrospect of the Great War*. New York: Macmillan, 1923.
55. Menninger, W. C. Psychiatry and the army. *Bull. Menninger Clin.*, 1944, **8**, 89–98.
56. Menninger, W. C. Perspective of psychiatry. *Ann. intern. Med.*, 1945, **22**, 170–181.
57. Menninger, W. C. Modern concepts of war neurosis. *Bull. Menninger Clin.*, 1946, **10**, 196–209.
58. Miller, E. Psychiatric casualties among officers and men from Normandy. *Lancet*, 1945, **248**, 364–366.
59. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
60. Mittelmann, B. Psychosomatics. In: P. L. Harriman (Ed.), *Encyclopedia of psychology*. New York: Philosophical Library, 1946, pp. 678–722.
61. Mowrer, O. H. An experimental analogue of 'regression' with incidental observation on 'reaction formation.' *J. abnorm. soc. Psychol.*, 1940, **35**, 56–87.
62. Mowrer, O. H., & Viek, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, **43**, 193–200.
63. Murphy, G. *Personality*. New York: Harper & Bros., 1947.
64. Murray, H. A. *Explorations in personality*. New York: Oxford Univ. Press, 1938.
65. Murray, J. M. Psychiatry in the Army Air Forces. *Amer. J. Psychiat.*, 1943, **100**, 21–24.
66. Myers, H. H., & Von Koch, S. Reactive depression: a study 100 consecutive cases. *War Med.*, Chicago, 1945, **8**, 358–364.
67. Needles, W. Statistical study of 100 neuropsychiatric casualties from the Normandy campaign. *Amer. J. Psychiat.*, 1945, **102**, 214–221.
68. OSS Assessment Staff. *Assessment of men*. New York: Rinehart, 1948.
69. Patton, G. S., Jr. General Patton's war letters. *Atlantic Monthly*, 1947, **180**, 57–61.
70. Phillips, E. L. War neurosis. II: A second study of clinical and control groups. *J. clin. Psychol.*, 1946, **2**, 155–164.
71. Phillips, T. R. Leader and led. *Coast Artill. J.*, 1939, **82**, 48–57.
72. Procter, T. H. The motives of the soldier. *Int. J. Ethics*, 1920, **31**, 26–50.
73. Raines, G. N., & Kolb, L. C. Combat fatigue and war neuroses. *Nav. med. Bull.*, Wash., 1943, **41**, 1299–1309.
74. Rapaport, D. *Emotions and memory*. Baltimore: Williams & Wilkins, 1942.
75. Roethlisberger, F. J., & Dickson, W. J. *Management and the worker*. Cambridge: Harvard Univ. Press, 1939.
76. Rorschach, H. *Psychodiagnostics*. (Trans. by P. Lemkau & B. Kronenberg.) New York: Grune & Stratton, 1942.
77. Rosenberg, S. J., & Lanbert, R. H. Analysis of certain factors in the histories of 200 soldiers discharged from the army for neuropsychiatric disabilities. *Amer. J. Psychiat.*, 1942, **99**, 164–167.
78. Ruesch, J., et al. *Chronic disease and psychological invalidism: a psychosomatic study*. New York: Amer. Soc., Res. Psychosom. Problems, Inc., 1946.
79. Sargent, W., & Slater, E. Acute war neuroses. *Lancet*, 1940, **239**, 1–2.
80. Saul, L. J. Physiological effects of emotional tension. In: J. McV. Hunt (Ed.), *Personality and the behavior disorders*. (Vol. I.) New York: Ronald Press, 1944, pp. 269–305.

81. Saul, L. J. *Emotional maturity.* Philadelphia: Lippincott, 1947.
82. Schreiber, J. Morale aspect of military mental hygiene. *Dis. nerv. Syst.*, 1943, **4,** 197–201.
83. Schwab, S. Acute war neuroses. *New Engl. J. Med.*, 1940, **223,** 431–433.
84. Sears, R. R. Survey of objective studies of psychoanalytic concepts. *Soc. Sci. Res. Coun. Bull.*, 1943, No. 51.
85. Sheldon, W. H., & Stevens, S. S. *The varieties of temperament.* New York: Harper, 1942.
86. Sheps, J. G., & Coburn, F. E. Psychiatric study of 100 battle veterans. *War Med.*, Chicago, 1945, **8,** 235–237.
87. Sherrington, C. *The brain and its mechanisms.* New York: Macmillan, 1933.
88. Simon, A., & Hagan, M. Social data in psychiatric casualties in the armed services. *Amer. J. Psychiat.*, 1942, **99,** 349–353.
89. Steckel, H. A. The duties of a consultant on neuropsychiatry on the staff of the chief surgeon. *Milit. Surg.*, 1936, **81,** 119–123.
90. Steinberg, K. L., & Wittman, M. P. Etiological factors in the adjustment of men in the armed forces. *War Med.*, Chicago, 1943, **4,** 129–139.
91. Stevenson, G. S. The prevention of personality disorders. In: J. McV. Hunt (Ed.), *Personality and the behavior disorders.* (Vol. II.) New York: Ronald Press, 1944, pp. 1164–1192.
92. Strecker, E. A., & Appel, K. E. Morale. *Amer. J. Psychiat.*, 1942, **99,** 159–163.
93. Strickland, B. A. The importance of psychiatry in military medicine. *Milit. Surg.*, 1938, **83,** 83–88.
94. Sutton, D. G. Naval psychiatric problems. *Amer. J. Psychiat.*, 1940, **97,** 255–275.
95. Torrie, A. Psychosomatic casualties in the Middle East. *Lancet*, 1944, **246,** 139–143.
96. Weiss, E., & English, O. S. *Psychosomatic medicine: the clinical application of psychopathology to general medical problems.* Philadelphia: Saunders, 1943.
97. Wexler, M. Measures of personal adjustment. In: D. B. Stuit (Ed.) *Personnel research and test development in the Bureau of Naval Personnel.* Princeton: Princeton Univ. Press, 1947, pp. 126–174.
98. Will, O. A., Jr. Psychosis in naval inductees with less than fifteen days of active duty. *Nav. med. Bull.*, Wash., 1944, **43,** 909–921.
99. Will, O. A., Jr. Certain psychiatric reactions in operators of antisubmarine sound gear. *Nav. med. Bull.*, Wash., 1945, **44,** 746–748.
100. Witherspoon, M. M. Morale in the new World War. *Proc. U. S. nav. Inst.*, 1940, **66,** 675–684.

Chapter 22

MORALE AND LEADERSHIP

ROBERT L. FRENCH

Northwestern University

Introduction

Other chapters of this volume deal primarily with two types of problems: the design of material facilities for efficient human use, and the design, so to speak, of human individuals for most efficient use of these facilities. The problem of morale and leadership concerns the effectiveness with which these individuals employ their resources when brought together into a functioning group—submarine crew, office force, training center staff, research unit, or whatever the case may be. Obviously these various problems are not clearly distinct; all relate to the same ultimate objective, and to a considerable extent all demand the application of similar concepts and methods. However, the study of functioning groups requires attention not simply to the actions or potential actions of individuals considered as isolated units, but also, and primarily, to the complex forms of interaction between individuals which characterize a group. In this respect the study of problems of morale and leadership probably calls for supplementary concepts and methods, perhaps quite different in nature from those appropriate in individual psychology.

Despite an enormous literature dealing with such topics as morale and leadership, the actual scientific study of group processes is of relatively recent origin. Hence, the necessary methodological and conceptual tools are still in an embryonic stage, and our knowledge of group functioning is accordingly meager. This chapter cannot, therefore, present a record of accomplishment and define precisely the steps to be taken next. Nor will it attempt simply to review the literature. Rather, an effort will be made to outline what seem to be the major problems, and within this framework to suggest on the basis of relevant findings and available techniques some further questions that might be asked and approaches that might be taken.

The Problem of Group Effectiveness

Some Aspects of Institutional Groups

The formal structure of a large institution such as a navy consists of many diverse and interrelated groups. For our purposes, we may consider a group in the formal sense to be any aggregate of two or more individuals which is regarded administratively as a unit, whether it be a fleet or a maneuvering room watch. Within an institutional organization, groups may vary in many ways. Certain of their characteristics are defined by orders from superior authority. These commonly include: general objectives; size; formal structural differentiation with respect to sub-groupings and individual roles, responsibilities and authority; formal relationships to other groups within the organization; general provisions for admission or release of members; relatively permanent material equipment (e.g., a submarine); and so on. A number of groups within the organization may be highly similar in their formal characteristics (e.g., submarine crews), in which case we may speak of a "type" of group.

Certain other group characteristics are less amenable to official regulation, and with respect to these, groups even of the same formal type may vary widely. Such characteristics involve the personal characteristics of members, the relatively stable informal interrelationships which they develop within the context of formal structure, such as patterns of leader-follower interaction and informal work groups, and so on. These informal characteristics are not always sharply distinguishable from the formal ones, but to make the distinction is useful for certain purposes.

Definition of the Problem

Broadly speaking, we may conceive the effectiveness of group performance with reference to group goals to be a function of group characteristics and factors external to the group. In this connection, formal characteristics are doubtless quite as important as the others, and the overall study of group effectiveness would have to treat these as unknowns in raising the general question: given certain goals to be attained, what is the best social group for the job? Doubtless this question will one day be tackled scientifically, but at present it involves too many unknowns. In the present discussion, the emphasis will be upon the role of "informal" characteristics in group functioning, and formal characteristics will for the most part be accepted as given. It is, however, necessary to consider the very real possibility that the influence of particular factors upon group effectiveness may vary depending on the formal characteristics of the group.

It seems necessary to limit the problem also with reference to certain other factors. Effectiveness in the attainment of a group goal, measured in objective terms such as the amount of tonnage sunk by a submarine, depends not only upon the characteristics of the group but also upon many relatively unpredictable external factors, such as weather, availability of targets, and so on. We are interested in the relative effectiveness of groups of a given type when these extraneous factors are held constant, or in what under non-laboratory conditions might better be called potential effectiveness. At the same time, these factors may influence subsequent potential effectiveness, and must therefore be considered as potential sources of variation in group performance. Similar considerations hold in the case of differences among groups of a particular type with respect to special material facilities or special technical skills possessed by members.

The general problem might then be stated as follows: given groups of a particular type with comparable material facilities and facing comparable external obstacles, how can their effectiveness be evaluated, what factors are related to effectiveness, and how can these factors be manipulated to advantage?

Some Fundamental Assumptions Concerning Group Behavior

To orient the subsequent discussion it may be well to review briefly some basic assumptions concerning group behavior which seem to be at least implicit in much of current thinking among social psychologists.[1] These by no means qualify as systematic theory, but they do point to general factors of probable importance for the problem.

First of all, it may be assumed that individuals belong to a group and work effectively in it because they perceive doing so as a means to satisfaction of their needs. The dominant needs may be complex in a given individual and may vary from one individual to another, involving perhaps, needs for status, companionship, money, satisfaction of ideals, domination of others, fear of punishment for not taking part, etc. Each individual, moreover, will have certain additional needs which group activity must not threaten if he is to participate without internal conflict. Or, conversely, the greater the extent to which group activity satisfies or promises to satisfy needs of any sort, the more effective, presumably, the individual's participation. For effective group performance, too, it is necessary not simply that important needs be satisfied in the group, but that individuals perceive their own partici-

[1] A good general discussion along these lines is presented by Krech and Crutchfield (**61**).

pation in collective action toward group goals as essential to satisfaction. Thus, an effective group should include individuals who have adequately strong needs, not necessarily the same in nature but alike in that their satisfaction demands the attainment of group goals without threatening the satisfaction of other needs, either their own or those of others. Conditions within the group or external to it which affect the prospects of need satisfaction may be expected to affect group performance.

Secondly, what is perceived as a means or as a threat to need satisfaction depends upon a number of factors. These include not only the "objective" situation as independent observers might describe it, but also various factors pertaining to the individuals involved: their needs; their levels of aspiration and the way they structure their goals in time; their beliefs and attitudes concerning human relationships, authority, proper ways of doing things, and so on. In general, probably, the more similar the characteristics of group members in these respects, the more effectively the group functions. And events within or without the group which influence these bases of perception may be expected to influence group functioning.

Thirdly, effective individual participation in group activity demands appropriate skills. Members of a group must be able to communicate effectively, which means that they must have in common not only a knowledge of the same language but probably certain other characteristics as well. In addition, many group activities require essentially similar skills of other sorts on the part of most or all members. At the same time, all groups involve a division of labor, partially recognized in the formal structure, partially developing informally through the interaction of particular individuals. A satisfactory distribution of skills appropriate to the varying functional roles within a group is thus important also for group performance. And again, situational factors affecting relevant individual skills must be taken into account.

Group functioning is a matter of individuals functioning in relationship to one another. Such concepts as need, belief, and skill, refer to the behavior potentialities of individuals. These may be conceived as underlying the system of interpersonal relationships which comprises the group. In some respects, as implied in the preceding paragraphs, it is feasible to describe a group using aggregate statements in terms of such concepts, e.g., that the members want or believe certain things. But to take account of all of the important and complex patterns of relationships within a group would seem to require descriptive schemes which go beyond mere aggregate statements, and which, for purposes of economy, involve concepts of a more inclusive order than those discussed above. At present well established concepts of this sort are lacking, though efforts are being made in various quarters to develop them (**43, 61, 65**).

It is perhaps unnecessary to add that group characteristics may change in consequence of group experiences, and that individual potentialities may change in the process. The full story of group functioning and of the role of individual behavior in it must give such changes a prominent place.

Morale and Group Effectiveness

The Concept of Morale

The abundant literature on morale contains many and varied definitions of the term, but in those definitions which imply any operations for observing morale (and the number of such definitions is considerably less than the total) there seems to be tacit agreement that the term denotes aspects of group behavior indicative of group effectiveness. This at any rate is the general conception that will be entertained here. Thus, if a group of a given type achieves its goals more efficiently than another equipped with similar technical facilities and facing equivalent external difficulties, the former may be said to have higher morale. If it can be shown that groups which achieve their goals efficiently exhibit a high degree of cohesiveness, think well of their leaders, do not fight much among themselves, agree on their objectives, have confidence in their equipment, and so on, then these manifestations represent high morale; but only if a relationship to goal attainment can be shown. Individual happiness, enthusiasm, "good" behavior, and so on, are all very desirable, but from the standpoint of group functioning they are irrelevant unless they indicate the presence of conditions conducive to the attainment of group goals.

Thus, group morale may be considered operationally equivalent to group effectiveness, or potential effectiveness, as it has been defined here.[2] But this is only a beginning. What is needed for research purposes is a more detailed and precise behavioral definition. In working out such a definition, it would seem that extensive empirical observation and analysis might well supplement the worn armchair favored by the majority of past writers on morale. Starting from general assumptions such as those previously discussed, it is possible to elaborate various hypotheses as to phenomena which should appear in groups functioning at various levels of effectiveness, e.g., that in poorly functioning groups the frustration of important needs of members should produce a relatively high level of aggression. It would then be desirable to secure for a series of groups of a given type as richly varied an assortment of measures of these phenomena as possible. As will be noted later, there are many possible measures of morale which seem promising. The interrelationships of these measures might then be analyzed to ascertain the possible existence of clusters or factors showing relatively independent variation. Since the possible manifestations of effective functioning are numerous and the underlying conditions complex, it is not too much to suppose, as many writers have suggested, that morale may comprise a number of different variables. Application of factor analysis procedures to data of the above sort would serve at least to clean out some of the underbrush and provide a conveniently economical, although tentative, behavioral definition.

Parts of this necessary empirical program have attracted the attention of investigators. Many have hypothesized sets of morale components or categories of morale phenomena. A few (e.g., **74, 42, 29**) have developed measures on the basis of such hypotheses, and made some analysis of the interrelationships in their results. One (**42**) has carried through a factor analysis. McNemar (**72**) has examined these studies critically with reference to techniques of measurement and analysis, and has suggested possible improvements. Of broader significance, however, is the preoccupation of these investigations with questionnaire data and with populations of individuals.[3] Further work aimed at clarifying the concept of morale could fruitfully employ a greater variety of measures, and follow the suggestion of Cattell (**17**) in applying factor analysis to a population of groups.

Methods of Appraising Morale or Group Effectiveness

A number of social psychological techniques are adaptable to obtaining the kind of

[2] In a discussion of factors influencing combat effectiveness in air groups, Darley (**26**) has raised the question as to whether morale may not be considered equivalent to psychological effectiveness. To the present writer this seems the most fruitful working assumption.

[3] It is, of course, possible to develop a concept of individual morale, with reference either to an individual's life in general or to his membership in a particular group (**21**), and the term is often employed popularly with both of these connotations, but in the evaluation of group functioning the group is properly taken as a unit. It will be seen in the discussion of possible measures of group morale that some have no meaning in terms of specific individuals. In the case of other measures, which involve gathering primary data on an individual basis, a rough measure for the group may be obtained through some process of averaging.

data needed in the study of morale. The majority of these have been discussed in an excellent review by Child (**21**). Here it is sufficient to review them briefly and indicate certain possible lines of further development.

The Use of Questionnaires

As was noted above, most of the studies of morale reported in the literature have employed questionnaires or attitude scales. This has been true in academic investigations of civilian morale (**74, 42, 29**), in most industrial morale surveys (**13**), and in the extensive studies of Army morale by the Research Branch of the Information and Education Division. In general these questionnaires have used fairly direct questions or statements relating to satisfaction with various aspects of group activity. They have varied in other respects, such as the degree of topical reference shown by their contents, and the techniques and sophistication employed in their construction. The Guttman-type scales (**40**) employed by the Army for measuring specific opinions presumably indicative of morale probably represent the highest standard of precision attained in studies of this sort.

It is possible that research will disclose that pencil-and-paper techniques are by themselves sufficient to give a comprehensive picture of morale in a group; and this result would indeed be a welcome one. Unfortunately, the necessary research has not yet been carried out. It is not known, for example, how groups that complain on paper behave on the job. A somewhat more guarded attitude toward these seductive devices may therefore be in order.

Observation and Interview Methods

Child (**21**) has stressed the probable fruitfulness of a more clinical approach to the study of morale, utilizing direct observation and non-directive interviews of the sort used so effectively by Roethlisberger and Dickson (**83**). Relevant objects of observation or inquiry might include: the nature and extent of interpersonal contacts; characteristics of individual or group performance on the job; verbal expressions in interviews, conversations, or in writing, of opinions concerning the group, reactions to outside influences, rumors, and the like; leisure-time activities; behavior in audience situations; and so on. The actual choice of behavior categories for systematic investigation would, of course, depend upon the hypotheses adopted at the outset. Various techniques are available or in process of development for quantifying such observations through the use of ratings, time-sampling, content analysis, etc. (See, for example, **4, 6, 19**).

Sociometric Procedures

The sociometric procedures of Moreno (**75, 30**), utilizing data obtained through questionnaires, observation or interviewing, provide a simple and illuminating picture of interpersonal relationships within a group. An interesting application of this technique to aircraft carrier pilot groups has been reported by Jenkins (**53**). Several methods have been suggested for deriving from such data a single measure of cohesiveness in a group (**98, 25**). Advances in sociometric analysis which will take into account qualitative differences in the bases for interpersonal choices may yield even more fruitful information concerning the level of group functioning.

Projective Techniques

Recent adaptations of pictorial projective techniques to the study of individual attitudes toward various groups (**81, 15**) suggest the possibility of extending such techniques to the study of morale. The individual's view of his own group situation might also be investigated through adaptations of verbal projective devices, such as the sentence completion test employed by the OSS assessment staff (**79**) in the study of individual personality. A technique recently developed at the Research Center for Group Dynamics requires a small group, as a group, to evolve a story about one or more test pictures; the story and the protocols of its development are reported on the basis of preliminary trials to tell much concerning the functioning of the group.[4] Another potentially valuable projective device applicable to small groups is the sociodramatic technique of Moreno (**77**).

Group Indices

Several writers (**1, 13**) have called attention to various kinds of indices that might be obtained from group records covering a period of time. In a military situation these might include statistics on disciplinary offenses, men on "sick call," psychoneurotic and psychosomatic disorders, venereal diseases or other diseases such as malaria in which incidence may depend on the taking of routine precautions by the men, requests for transfer, men striking for higher rates, and so on. There are obvious problems in the use of such indices: they can give only a kind of average picture for a period of time, in practice the necessary records may be incomplete or unavailable, and they may be influenced unduly by extraneous factors. Nevertheless, further exploration in this direction may be worthwhile.

Situation Tests

Additional evidence of a more dynamic sort might be gained by observing group reactions in standardized stress situations. In effect, whole groups might be subjected to situation tests analogous to those recently introduced in studies of individuals as potential leaders. Some suggestive possibilities may be found in the literature, e.g., a simulated fire with a group in a locked room (**32**), intrusion of a critical stranger (**66**), false news announcements (**78**), together with various situations of a more conventional sort, such as group problems, inter-group competitions, etc. Measures of aggression, fear, persistence, acceptance of destructive rumors, etc., in such situations might prove very revealing of a group's potentialities. While procedures of this sort would perhaps have their greatest usefulness in relatively small scale research investigations, it is not unlikely that some realistic test situations could be devised for more or less routine application in conjunction with regular group training programs, e.g., during shakedown, or that pertinent data might be gathered in stress situations which already occupy a place in such training.

The Validity of Morale Measures

It was pointed out earlier that direct measures of the success of groups in attaining their goals may be influenced heavily by adventitious factors, and that our problem concerns the relative effectiveness of different groups when differences in these factors are, at least conceptually, eliminated. Actually to eliminate such influences is extremely difficult (**13**), and for that matter, relevant direct measures may in practice be unobtainable. How then can it be shown, as we have also demanded, that potential measures of morale or group effectiveness are valid?

The answer, it would seem, is that we must in some degree lift ourselves by our boot straps. Certain potential measures have a high degree of face validity. If a group spends most of its time in wrangling or in loafing, or has a high proportion of deserters, it obviously cannot be functioning very effectively. Whether or not such measures are available, an analysis of interrelationships among measures may also yield relevant evidence. If several morale indicators are hypothesized on the basis of related fundamental assumptions, and, when measured, are found to be highly intercorrelated, this should justify at least a tentative presumption of validity. Likewise, if certain effects on morale are predicted to follow certain events, and a presumed morale indicator shows the predicted effects, this may be considered relevant evidence.

Further evidence on validity may be obtained in experimental group situations affording control of extraneous influences on

[4] Reported in *The TAT Newsletter*, 1948, **1**, No. 3, edited by R. R. Holt, Menninger Foundation, Topeka, Kansas.

performance. Results obtained in such situations are likely to have doubtful generality, but they may suggest promising leads.

Factors Related to Group Effectiveness

Theoretical considerations and common experience point to the probable importance of various factors for morale or group effectiveness. These are classified here under three heads: characteristics of group members, conditions of group living, and leader-follower relationships (discussed in a separate section). These categories obviously overlap, and in their effects upon group functioning the various factors interact with one another in a complex fashion. To a considerable extent, moreover, the relationship to group effectiveness is circular; the factors discussed are likely to be both symptoms and causes of a given level of group effectiveness. Within rather broad limits, however, it would seem to lie within the power of authorities at one or another level in an organization either to manipulate these factors to advantage, or to anticipate their effects and take appropriate action.

Characteristics of Group Members

Minimum Prerequisites for Participation

Certain individual characteristics involved in learning skills and in getting along with people are necessary in all group members if any group within an organization is to function effectively. In military organizations the rejection or special assignment of individuals deficient in these respects is the job of medical, neuropsychiatric, psychological, and security screening agencies. These problems lie outside the scope of this chapter, but we may note their relevance.

Distribution of Personality Characteristics Within a Group

If group functioning involves individuals functioning in relationship to one another, then the characteristics of individuals assigned to a group should be considered not simply with reference to their satisfying certain common minimum standards, but also in terms of the way they fit together. Ordinarily attention is given only to the distribution of special skills needed to fill the roles defined in the formal structure of the group. But conceivably the patterning of other characteristics may be of equal or greater importance. In a paper written for a class assignment one of the writer's students commented spontaneously on "the Army's haphazard, lackadaisical method of combining ten individuals into an aerial combat team, without the slightest regard to personality, interests, physical appearance, age or other characteristics of the crew members. I have seen much strife and occasionally bloodshed among crew members as a result."

This problem raises some interesting and methodologically somewhat complex questions which research has scarcely touched. Do groups function best when they are composed of individuals of similar characteristics? Common experience would suggest that with respect to at least certain characteristics, such as dominant interests and values, this may be the case. In the case of married couples, some evidence (**90**) supports this supposition. An American naval observer (**96**) wrote in 1940 that in the German Army congenial men of similar interests were placed in the same units, and that the High Command considered the results "astonishing." Whether this policy continued throughout the war is not known, but it seems doubtful.

On the other hand, may there not be other characteristics with respect to which some variability is desirable in a group? One might suppose, for example, that a group composed entirely of rather insecure, dependent individuals would not function well, and that neither would one consisting of highly aggressive, dominating individuals, but that a combination of the two might do quite adequately. Again, some evidence on married couples (**90**) is suggestive, but more data are needed.

Some interesting evidence relative to optimal group composition comes from sociometric studies in which individuals select one another as members of a group. For example, at the New York State Training School for Girls, Jennings (**55**) compared the adjustment of new girls assigned at random to living cottages with that of newcomers assigned on the basis of mutual selection by the girl herself and by leaders in the cottages. The results showed better adjustment for cases of mutual selection. Similar findings have been reported by other investigators (**76, 58**). Zeleny (**99**) has reported preliminary studies involving sociometric assignment of pairs of Army cadet pilot-observers, but without evidence of the relative effectiveness of this procedure. The role of personality factors in interpersonal affinities of this sort is not well understood. Considerable attention has been devoted to possible personality differences between highly selected and unselected individuals, but none, as far as the writer knows, to the personality patterns which characterize a self-selected group. A further question of potential practical import concerns the effects of the choice process itself. Would equally good results be achieved if persons who would choose each other if given the opportunity were assigned arbitrarily to the same group?

Job Assignment

Related to the problem of patterning of characteristics is that of the adequacy of assignment of individuals to technical roles within the group. This is presumably important not only in terms of filling each job with the best qualified man, but also in terms of the frustration which may arise if men cannot employ their best and most interesting skills. The Research Branch of the Army Information and Education Division accumulated considerable evidence to suggest that "satisfaction with job assignment is perhaps the single most important factor of morale" (**45**). Thus in a survey made at the end of the war, 67% of soldiers did *not* feel that in most ways they had gotten a square deal from the Army, and among these men, job assignment was the most frequent source of spontaneous complaint (**49**). An earlier study (**47**) showed that 74% of men who had gotten the job assignment they had asked for expressed a high degree of job satisfaction, whereas among men who had not gotten the assignment asked for, or who had had no chance to ask, 19% expressed high satisfaction. Since these studies depend upon statements by the men after assignment, there is no possibility of determining the extent to which other personal factors related to morale may have accounted for failure to be offered a choice or to be given a chosen assignment. The actual effects could be decided definitely only by an experimental procedure. It is probably true, also, that other factors may compensate for failure to obtain preferred assignments, and conversely, that job assignment may serve as a prominent hook on which to hang dissatisfaction arising primarily from other sources. Despite these qualifications, it is highly probable that this is one aspect of personnel policy of considerable importance for morale.

Training

Since individuals rarely exhibit at the outset all of the characteristics necessary for effective contribution to military groups, it is necessary to modify their characteristics through formal training procedures. Technical training represents one aspect of this problem. This is discussed elsewhere in this book.

It is also desirable, however, that members of military groups have in common many general habits, beliefs and attitudes relative to the organization, its objectives, its customs, and so on. Recruit training, orientation training and the like are presumably designed to expedite the acquisition of these. There appears to be little evidence concerning the contribution of established procedures of this sort to the effectiveness of military groups. In connection with the Army Orientation Program, the Research Branch conducted extensive experimental studies of the effectiveness of films and discussions dealing with the broader phases of the war, and found generally some changes in specific opinions. To what extent these changes were accompanied by changes in other aspects of group behavior is not known.

Various devices, used both in connection with such training programs and as a part of regular military life, have long been believed to encourage the identification of individuals with military groups, and thereby to contribute to group effectiveness. Some of these are: uniforms, insignia, colors, the numerous rituals involved in interpersonal relations and mass observances, devices serving to accent the identity of particular groups, and mass activities such as group singing and close-order drills. There appears to be no precise evidence bearing on the effectiveness of these procedures, and it may be that none is necessary. However, traditional practices are liable to reinterpretation under changed conditions, and it is conceivable that an activity such as close-order drill, if perceived by participants as an outmoded device useful only for killing time, would defeat its purpose. Possibly experimental evaluation would be in order in some instances.

Length of Time Spent in the Group

A study carried on in the British Army during World War II is reported to have shown that incidence of various psychological and medical disorders reaches a peak after about three months of service overseas and thereafter tends to subside.[5] Various studies of group behavior in other contexts have taken notice of changes with time. Benge and Capell (**11**) have reported a regular relation between industrial morale, measured by questionnaire, and length of individual service. Jennings (**56**) and Bronfenbrenner (**14**), to cite but two investigators, found fewer isolates and, in general, greater solidarity in newly formed groups of young people after periods of months. Obviously the effective variable in these cases, whether individual or group, is not sheer duration but what happens in time. However, where time can be shown to be related to indices of morale, then at least these fluctuations can be anticipated and allowed for, and possibly in some cases conditions responsible for low points can be isolated and remedied.

Personnel Changes in a Group

Many considerations of military policy render flexibility in shifting personnel desirable. On the other hand this has been alleged to interfere with effectiveness, particularly in small, tightly-knit units such as bomber crews. There is no conclusive evidence on this point. Ward (**93**) reported that Army bomber crews which had had changes of bombardier or pilot showed larger combat strike errors than crews without such changes. He pointed out, however, that the crews had not been equated in proficiency prior to the changes, and that personnel changes might be more likely in poor crews. The possibility that personnel changes may under certain circumstances have positive effects is suggested by the following observation in the ninth patrol report of the USS TAUTOG (**28**): "The policy of rotation is a most important morale factor. An enthusiastic new hand... and most of them are... acts as a stimulating tonic. Everyone takes delight in demonstrating their knowledge and feels more keenly their responsibility." This is clearly a question calling for experimental analysis, in groups at various stages of development

[5] Described to the writer by Dr. A. T. M. Wilson.

and with reference to changes in various classes of personnel.

Conditions of Group Living

General Conditions Within the Group

Among conditions recognized as important to individual satisfaction and effective performance in military as in other group situations are: financial incentives, housing, clothing, sanitary and medical facilities, food, sexual satisfaction, recreational facilities, mail delivery, schedules of work and rest, furloughs and passes, quality of equipment for working and fighting, opportunities for promotion, and so on. Unquestionably, severe deprivation in any one of these respects may have deleterious effects upon group functioning and morale. As with job assignment, however, this is probably not all of the story. On the one hand, complaints about such conditions may be heard in situations where impartial investigation discloses no particular deficiency. Thus, the Research Branch (**48**) found that WAC companies with unfavorable opinions of their officers had less favorable opinions of their physical facilities than did companies that thought well of their officers, even though no differences in such facilities could be detected upon investigation. Similar examples could be drawn from industrial relations. Thus, complaints may indicate a disease, but they do not necessarily diagnose it.

On the other hand, deprivations in any of these respects can probably be tolerated within rather broad limits depending on the total context of conditions. Within these limits threats to group effectiveness may arise, apparently, through the cumulative effect of a multitude of deprivations, or when morale is already low, or when deprivations are regarded as unnecessary, as unfair discrimination, as a result of incompetence on the part of officers, as threats to status or social relations, as products of a system which has no concern for individual welfare, and so on. These at least are reasonable hypotheses, supported by a variety of observations, such as those of Roethlisberger and Dickson (**83**) and of the Army Research Branch, and calling for more systematic investigation. They point also to the importance of the more obviously social factors in the group situation. Among these factors, suggested for the most part by case studies in industrial relations (**83, 38, 70**) are the following: a friendly attitude on the part of superiors, a consistent system of rewards and punishments, adequate basis for knowing what is expected of one and what part one plays in the "big picture," opportunities for adequate social relations with other members of the group, opportunities for participation in planning and for exercising initiative, and machinery for the appeal of grievances. McGregor (**70**) has suggested that the importance of some of these factors may depend upon the presence of others; thus, for example, men are not likely to take advantage of opportunities to show initiative unless they feel secure in their conditions of work and relations to superiors, but if the latter conditions exist, absence of the former may produce discontent. These considerations point in turn to the central significance of leader-follower relationships.

Relations to Other Groups

Any particular group functions in a matrix of other groups, both within and outside of the organization, and its relationships to them must inevitably influence its effectiveness. We know very little about such effects, but a number of obvious conjectures can be made which might deserve investigation.

It seems reasonable to suppose that a large measure of an individual's satisfaction with, and pride in, his own group may derive from such relationships. For one thing, features of one's own group—traditions, living circumstances, successes, leadership—tend to be judged relative to those believed to exist in other groups. At the same time, the opinions which other people hold toward a group may influence the opinions of the group toward itself.

Competition between groups may spur effectiveness. Also, where cooperation between groups is necessary to effective performance, morale may be involved. The failure of another group on which success or safety depends, or conditions creating lack of confidence in it, may contribute to low morale. Lack of confidence in supporting groups may also serve, of course, to indicate low morale arising from other causes: witness the scapegoating of rear echelons by front echelons, of civilians by military groups, and vice-versa. Scapegoating may function to preserve unity in a group, but when its targets are cooperating groups or individuals, the overall effects may be undesirable. Investigation is needed of the conditions under which scapegoating can be directed to advantage in cases where conditions causing discontent cannot be remedied.

Additional influences on morale may stem from individuals' memberships in other groups, notably families. The effects of bad news from home on soldiers or sailors are proverbial. Such influences do not initially affect all members of a group, but, presumably depending upon the position of the individual involved, the effects may conceivably spread to other individuals. Studies of communication within groups would enlighten us on this and related points (e.g., spread of rumors) and perhaps suggest prophylactic measures.

Factors Peculiar to Combat

Considered in terms of psychological processes, there are perhaps no factors peculiar to combat; for some groups or individuals, at any rate, deprivation, danger, anxiety, may be extreme under quite "peaceful" circumstances. For others combat may provide a release from tension. But the tremendous increase in stress which combat entails in most cases poses additional, or different, problems in the practical control of morale.

Length of combat tour is undoubtedly a factor, but its relative importance must obviously depend upon the particular situation. When infantrymen overseas were asked by the Research Branch (**50**), "If you had a chance to change one thing in the Army, what would you change?", their most frequent response (**26%**) was to suggest rotating front and rear-echelon troops, setting a time limit on combat, etc. On the basis of interviews with naval fliers whose organizations had been in combat from seven to nine months, Darley (**26**) concluded that length of tour was much less important for combat effectiveness than some other factors, among them "adherence to a promised relief date." Quite probably the setting of a definite relief date and strict adherence to it would be important in all situations.

Intensity of combat, and number and distribution of losses, are among the more important factors mentioned in the study reported by Darley. Losses have the obvious effects not only of disrupting organization, especially if they involve key men, but of augmenting anxiety among men who remain. In some cases survivors may experience guilt over the fact that they have survived (**57**).[6] On the other hand, conceivably a few losses may, by heightening awareness of the seriousness of a situation, increase determination and effort. This would probably hold primarily for inexperienced units not thrust too suddenly into extremely dangerous combat.

Severe combat can also make for greater solidarity among members of a group, though this may or may not contribute to

[6] Child (**21**) has suggested that guilt consequent on killing the enemy must play an important part in morale, and this may be true of many individuals, at least transiently, but more commonly guilt seems to be attached to feelings of not having more dangerous jobs, of having let the outfit down, and so on.

overall effectiveness. A dramatic illustration is found in the report of the first war patrol of the USS PUFFER (**28**). Held down for a record 38 hours under more or less constant attack by Japanese patrol vessels, the crew reached a state of virtually complete exhaustion, and for some time after escaping showed marked signs of "nervousness." On subsequent patrols, however, it was observed that such strong bonds had developed among the men that new replacements who had not been through "the depth charging" were never fully accepted. The Commanding Officer recommended in consequence that crews surviving such an ordeal be broken up, presumably so as to allow the unrestricted operation of the usual personnel replacement policy. These various effects might be expected to depend somewhat on the characteristics of the situation and of the group involved; for example, the size and solidarity of the group, the extent to which individuals are equally exposed to danger, or the degree to which survival is an all-or-none proposition, should make a difference.

Additional factors important for morale in combat include the nature of duty between combat tours (**26**) and policies governing relief from combat on psychiatric or other grounds (**57**). Some of these aspects of the combat situation are discussed in another chapter in relation to individual effectiveness.

Group Success and Failure

That tired aphorism, "nothing succeeds like success," refers directly to a phenomenon of morale. Judging from Duff's (**28**) survey of war patrol reports, submarine commanding officers agreed universally that nothing contributed so greatly to morale as successful action, and that, conversely, morale was seriously undermined by long periods without adequate surface contacts. It may be that the effects of success or failure would not be so marked under less arduous or confining conditions, but this factor plays a vital part nevertheless in all individual and group behavior. Success operates, presumably, to confirm the validity of a particular approach to a goal, and thus to create confidence in oneself and in the group. A tradition of success appears to make for greater solidarity and for the more rapid assimilation of new members.

Success or failure are effective, however, only as they are experienced, and they are experienced only in relation to the goals perceived by members of a group. This, makes possible some control of this factor, through such measures as setting realistically high goals and sub-goals, providing evidence of the possibility of their attainment, insuring clear recognition of them by members of a group, and furnishing adequate experience of the results of action (**63**). Under some conditions, at least, participation of members of a group in setting a goal increases their awareness of and involvement in it (**10, 64**).

Just as in simple learning situations, however, failure may have educational value and make for a more securely based long-term morale. Probably this would be the case where morale was not already low, and where the reasons for failure were clearly perceived and circumstances permitted remedial action. An unbroken string of successes, particularly if they are easy, may breed over-confidence and a falling off of interest.

"Positive" vs. "Negative" Factors

Various writers (**68, 61**) have insisted that the state of morale produced by a positive common goal or positive interpersonal attractions among members of a group is somehow different from and better than that brought about through such "negative" factors as a dominating leader or a threat to the group. Thus, a group held together "negatively" may work as well as a "positive" group when the leader is present, but not when he is absent. There is undoubtedly much truth at the bottom of this distinction, but the relationships implied need working through. The terms "positive" and "negative" have a moralistic ring which suggests that they be abandoned in favor of more precise definitions of the factors thought to be at work. And the possibility that the effects of such factors may vary depending on the presence or absence of other factors, as well as on the dimension of morale studied, should be explored.

Problems of Investigation

From what has been said, it should be plain that we are dealing with complex relationships about which we know very little in any precise sense. Virtually all of the statements made above are more in the nature of hypotheses than of established findings. Certain general problems confronting investigation might be noted.

First of all, research is handicapped by lack of an adequate concept of morale. The studies cited have employed various aspects of behavior as morale measures. While these may all be in line with the popular conception of morale, there is no assurance that they represent the same thing psychologically. The multiple dimensions of morale which will doubtless be discovered on inquiry will not react in the same manner to most influences.

Further problems concern methods of investigation.[7] It is apparent that many, perhaps all, of the conditions influencing morale cannot be treated in isolation—that morale is a function of interacting factors. For example, it was suggested that the effects of certain conditions might depend upon whether morale in general was already high or low, or whether due to the effects of other factors these conditions were perceived in one way rather than another. These considerations would argue for the use of factorial designs wherever possible in research, and for the limiting of conclusions to the range of conditions actually studied.

Due to the difficulties involved in applying strict experimental procedures in many practical situations, frequent use has been made of *ex post facto* procedures. A case in point is the Research Branch study of job assignment in relation to job satisfaction (**45**, discussed above). As in that case, the possibility frequently exists that other factors related to the dependent variable may have accounted for the grouping of subjects with respect to the presumably independent variable, and doubts on this question can often be resolved only by use of conventional experimental procedures involving prior design and randomization. In many cases the latter type of procedure can be employed, of course, and where this is true, the greater certainty which it provides may easily warrant the additional trouble involved.

Finally, it should perhaps be stressed again that simply asking people what is good or bad about their situation may indicate that *something* is good or bad, but it will not necessarily tell what it is. This phenomenon is so familiar to clinicians and so obvious in the literature of industrial relations as to need no emphasis, but morale investigators have sometimes shown a surprising trust in the insight and diagnostic abilities of their subjects.

Leadership in Relation to Group Effectiveness

The Concept of Leadership

Leadership has long been regarded as the most important factor in group effectiveness, and doubtless properly so, but a general conception of leadership conducive to fruitful research has been wanting. In many cases it has simply been confused with preeminence or with office-holding. In most cases it has been defined as a function of certain personality traits residing in the individual designated as leader.

[7] Various of these problems are discussed in greater detail by Kornhauser (**60**). Good general discussions of methodology are found in monographs by Greenwood (**39**) and Chapin (**18**).

More recent writers have advanced the conception of leadership as a functional relationship between individuals having different roles in a group.[8] This relationship is assumed to develop when there is a group goal important to the satisfaction of members' needs, and when following the suggestions of one member is perceived by the others as the most likely means to need satisfaction; that is, to attainment of the group goal with minimum frustration of other needs. It is thus a phenomenon dependent not merely upon the personal characteristics of the leader, but upon the characteristics of the group and of the external situation as well.

The functional character of leadership, and the complex interplay of these determining factors, may be observed especially clearly in small, informal groups (**12, 94**) or in situations of the "leaderless group" variety (**37, 79**), where a small, unorganized group is given a simple problem requiring coordinated effort with no instructions as to how to coordinate. In these situations it appears that leadership literally "emerges" as a product of interaction, and that it may change hands whenever the situation changes in a way that causes group members to perceive different suggestions or contributions as appropriate, and that calls forth such contributions from a different individual. It would seem, moreover, that in one situation or another almost any individual may, for a time at least, assume leadership.

In groups which form parts of large organizations the leadership process is complicated by virtue of the fact that leaders are normally appointed by superiors for relatively long periods. Thus, as Knickerbocker (**59**) points out, the leader is appointed as a means to the goals of his superiors, but to be followed, he must be perceived as a means to satisfaction of the needs of members of the group, or as a means to prevention of reduced satisfaction of needs. Moreover, the existence of his group in the context of a large organization implies relatively fixed expectations of the leader on the parts of superiors, equals, and followers; hence the leader's role and the process of leader-follower interaction may be heavily determined by the characteristics of the organization. Finally, the group situation changes in time, and the roles which the leader is called upon to play may vary widely. If he is to remain accepted in the same sense as the leader of an informal group, he must presumably possess the varied assortment of personal qualifications necessary to effective contribution in these roles. He may, of course, derive some advantage here from his official status and from prestige acquired in his more effective performances.

With reference to the question of the relationship of leadership to group effectiveness, it would seem desirable that research first attempt to determine what patterns of leader-follower relationships characterize effective groups of any given type, and then investigate the conditions under which more effective leadership, and through it more effective group functioning can be obtained. Traditionally the problem of securing more effective leadership in groups has been envisaged as a matter of changing the characteristics of leaders, primarily through selection, and secondarily through training, and this is very probably the most promising approach. However, if effective leadership depends upon the situation as well as upon the characteristics of the leader, the possibility of achieving effective leadership through situational changes should not be overlooked. In any case, situational factors must be considered, and leadership must first be studied as it occurs in specific situations. Otherwise it will never be possible to evolve adequate hypotheses as to relevant personality characteristics in potential leaders, or appropriate methods of leadership training.

[8] Particularly good statements have been given by Knickerbocker (**59**), Gibb (**37**), Jennings (**56**), and Barnard (**7, 8**).

Effective Leader-Follower Relationships

The discussion of other factors affecting morale pointed to many aspects of leader-follower relationships having probable importance for effective group functioning. Here it may be well to consider the question of potentially fruitful approaches to an enlarged understanding of these relationships.

As in the case of other factors influencing morale, the determination of effective leadership patterns can be approached by empirical or *ex post facto* procedures, as in the studies by Jennings (**56**), or experimentally, as in the studies of Lippitt (**67**) and Lewin, Lippitt, and White (**66**). Both types of approach are useful, although in the latter case a possible question arises as to the adequacy of experimental leaders merely *acting* in predetermined ways, especially when they may have preconceptions as to the most effective method. At present, however, we know so little about leadership processes that extensive empirical observation would seem desirable. This can serve to develop hypotheses which may then be tested by more rigorous experimental means where feasible.

Another question concerns general methods of observation. The study of effective leadership may be carried out through direct observation of leader-follower interaction, as in the studies first mentioned, or circuitously, in terms of what group members say about leadership. Examples of the latter are found in the use by the Air Force (**24**) of anecdotes about leader behavior, and of less specific descriptions by various investigators (**92, 91, 43**). Reactions of group members are important to a full description, but they are not sufficient. Individuals may be biased, and all members even of a small group may overlook important aspects of what has happened (**52**). Direct observation encounters practical difficulties, especially in organized groups of adults, but these difficulties must somehow be overcome if rapid progress is to be made in the study of leadership behavior.

What specific concepts and techniques offer promise in the analysis of leader-follower interaction? A considerable amount of recent thinking revolves about broad patterns or types of leader-follower relationships, often represented in dichotomized form, such as democratic vs. authoritarian (**66, 67**) leadership vs. domination (**80**), or integrative vs. dominative behavior (**2**). Along somewhat similar lines, Knickerbocker (**59**) has discussed four general patterns which are prevalent or possible in industry—force, paternalism, bargaining, and mutual means—and has suggested certain hypotheses to account for their apparent relative effectiveness. Basically he assumes—and the same assumptions seem implicit in the above writers—that maximum productivity is secured where attainment of organizational objectives is mutually perceived by followers and leaders as means to the satisfaction of their respective needs, and that the problem for the leader is to create conditions which make this possible. These considerations afford a general framework for investigation of leadership, but if analysis in terms of types or broad patterns is to be useful, they need considerably more explicit definition and on a more definitely empirical basis.

Another line of inquiry which appears highly promising involves the empirical analysis of informal roles and channels of communication to and from the leader. For example, Benne and Sheats (**12**) have classified and discussed the various roles which the nominal leader and other members may assume from time to time in discussion groups. A somewhat different but related approach has been reported by Stogdill and Shartle (**87**), who attempt in Naval groups to analyze the leader's interpersonal relations in terms of such variables as work patterns, personal contacts, delegation of authority, and methods of working with staff. The problem of informal communication systems developing within industrial organizations has attracted the attention of various investigators, among them Roethlisberger and Dickson (**83**), and Arensberg and McGregor (**3**). The latter, in a study of an electrical instrument company, determined the "structure of habitual relations" through interview and questionnaire methods, and found it very different from the formal organization chart. The low morale of design engineers, who constituted the major concern of the study, appeared to arise from their failure to accomplish anything through formal channels. The application of techniques of this sort to groups of similar type which differ in their level of effectiveness should add much to our knowledge of effective leader-follower relationships.

Still another potentially effective approach might involve the application of clinical methods to the analysis of affective relationships which develop between leaders and followers in groups varying in effectiveness. Various theorists (**35, 73, 88, 23**) have discussed this general question in various terms with reference to interpersonal relationships in general, but no systematic empirical studies have been reported.[9]

If meaningful dimensions of morale can be established through correlational and experimental analysis, as previously suggested, they might serve as a useful framework within which to analyze leadership patterns through any or all of the above approaches.

A final point may be noted here concerning the role of the leader in initiating effective relationships with followers. It has been assumed that followers contribute to the shaping of such relationships as well as do leaders, even in the most authoritarian groups. If this is so, then it is not sufficient simply to determine what forms of relationships exist in effective groups. Specific attention must be given also to the ways in which and the conditions under which appointed leaders can initiate and further effective relationships in a purposeful manner. Studies of the development of leader-follower interaction in groups, of the effects of leader replacement, and the like, should be relevant in this connection.

[9] Studies are being undertaken by E. W. Eng and the writer aimed at developing methods for the analysis and classification of interpersonal relationships within groups, including leader-follower relationships.

Effective Leadership in Different Situations

It is a commonplace statement in current literature that "leadership is relative to the situation." What is presumably meant by this is that the kinds of leadership behavior which develop, and the effectiveness of any given pattern of leadership, depend upon the formal and informal characteristics of a group and the external factors affecting it at a particular time. A given group may be thought of as involving a series of situations, or different groups as involving different series. As we have noted, there are good grounds for assuming that in both respects leadership is relative to the situation, but concerning the relationships implied we know relatively little. Most of the studies reaching the conclusion of relativity have been concerned with the general traits of leaders in different kinds of groups. (See Stogdill's review, **86**.) What need to be determined are the differences in effective leader-follower relationships which different situations demand. How do these interaction patterns vary in groups of different sizes, or of different degrees of homogeneity of members, or under varying conditions of stress? In what ways does the overall policy of an organization limit the freedom of a leader to initiate particular kinds of relationships with members of the group? These are but a few of the important questions which might be asked.

A significant exploratory study has been carried out by Hemphill (**43**), using questionnaire reports by members of almost 1000 groups as a basis for describing the groups in terms of 15 dimensions, and for describing the leader's behavior and evaluating its effectiveness. Hemphill found that variations in each of the group dimensions were associated with variations in the reported behavior of effective leaders. Further investigation of such relationships is obviously necessary, using methods of direct observation and taking into account additional situational differences, not only between but within groups. It is, however, too much to expect extensive development along these lines until more suitable concepts and methods have been developed for studying leadership in any particular situation.

The Selection and Training of Leaders

The problems of selecting and training leaders are, in principle at least, no different from those arising with various classes of technical personnel. The necessity for job analysis, development of reliable performance criteria, and the development and validation of selection and training techniques, is present in both cases. These general procedures are discussed elsewhere. Here it is sufficient to consider briefly some questions of their specific application to problems of leadership.

Job Analysis

Any particular leadership job is defined by the series of typical situations which it involves, that is by the formal characteristics of the group (including official specifications of the leader's duties), typical informal characteristics, and external situations typically encountered, as these are relevant to the leader himself. Now, the determination of the relevant aspects of situations is actually the objective of much of the needed research previously discussed, so that adequate analyses of leaders' jobs are at present difficult. And their significant features cannot be surmised with the same ease as in the case of a technician's job, which is defined to such a great extent by relatively inflexible physical equipment. Even now, however, probably considerably more can be done in the way of rough, tentative analysis and classification than has been done to date. Suggestions concerning the analysis of general leadership functions which supplement the preceding considerations of situational factors, have been made by Barnard (**7**) and Coffin (**22**).

Leadership Criteria

Measures of Group Effectiveness

A leader's effectiveness is measured by the contribution which he makes to group effectiveness. Thus, if other factors contributing to group effectiveness are equal for groups of a given type, the various measures of group effectiveness previously discussed may become criteria of leader proficiency. These other factors are rarely equal, but group measures may be valuable nevertheless, if treated with appropriate reservations. And, imperfect as they may be, they provide the ultimate criterion against which other criteria must at least in theory be validated.

In general it is desirable that criteria be analytic, in the sense of providing separate evaluations of performance on different aspects of a job as these are identified through job analysis. If meaningful morale dimensions can be used as a framework for analysis of leadership functions, a possibility considered earlier, then appropriate measures of group effectiveness might provide analytic criteria.

Judgments by Associates

Judgments of a leader by his associates have been used frequently in leadership studies, doubtless for obvious reasons. Questions arise concerning what associates to use and what kinds of judgments to ask of them. With respect to the first question it is a common practice to secure only the judgments of superiors. These are necessary, but not always sufficient. A leader must also work with subordinates, who often have quite different expectations, different bases for judgment, and, indeed, different opinions of him. When the Research Branch (**46**) asked, "What makes a good non-com?," 75% of commissioned officers rated the "ability to think for himself" as being first or second in importance, while only 23% of privates did likewise; at the same time 49% of privates rated "ability to gain the personal liking of the men" in first or second place, while only 7% of officers agreed. In general, it would seem desirable to elicit the judgments of subordinates and equals as well as those of superiors. Subordinates' opinions of a leader are, of course, often used as indices of morale.

Judgments may be secured in various forms: ratings of overall success or of success on different parts of a leader's job; ratings of relatively general traits assumed to be related to leadership effectiveness; check lists, ratings or other descriptions of more specific items of behavior believed to be necessary for effective leadership; man-to-man ratings; rankings or nominations of individuals preferred as leaders. The relative advantages and disadvantages of these techniques have been discussed by various students of personnel methods (e.g., **16**). Several points only need be noted here. Many such schemes appear to be analytic, in evaluating separate aspects of the job, behavior items, etc. Even if these categories are relevant to the leader's position, which is often doubtful, the judgments of associates may reflect only an overall evaluation. In general, rankings or nominations are simpler to administer and give as good if not better results than most of the others. The nominating technique has been used extensively in recent studies (e.g., **56, 82, 91**). It might be added that the forced-choice technique, embodied in the Army *Officer Evaluation Report, Form OER* and validated against a nominating criterion, seems to offer considerable advantages over conventional rating procedures (**85**).

Judgments by Independent Observers

Descriptions comparable in form to those of associates may be obtained through independent observers. If observers are adequately trained, and if circumstances permit systematic observation of the behavior of both leaders and followers, it is reasonable to expect evaluations having both higher validity and greater diagnostic value than the reports of associates alone.

Miscellaneous Criteria

Many studies of leadership have employed as criteria various facts relating to leadership status, e.g., mere holding of a position of leadership (as against not holding one), responsibility of position held, as measured in terms of rank, income, or some other factor, or number or frequency of past promotions. Doubtless these bits of information are generally related in some degree to performance on the job at hand, but they do not measure it directly, and the extent of the relationship obviously remains unknown.

In the military situation performance in Officer Candidate Schools has been a favorite criterion. Standing has been measured in various ways, such as by course grades, judgments by instructors or fellow students, etc. Here again, the value of such criteria obviously depends upon their correlation with performance on the job. Where it has been possible to determine this correlation, the results have not proved spectacular. For example, Leavitt and Adler (**62**) found correlations of Marine Corps OCS standing with a series of ratings by superior officers in combat ranging from −0.26 to 0.12 (while correlations of the combat ratings with sociometric status prior to OCS ranged from 0.04 to 0.44). It seems probable that school performance criteria can be improved in many cases, however.

As leadership theory develops and direct studies of organizational functioning increase, many new criteria should emerge. A possible example is the RAD Index developed by Stogdill and Shartle (**87**), though the full significance of this is not yet clear.

Methods of Selecting Leaders

Leader Personality and Effective Leadership

The problem of selecting leaders is one first of all of determining the characteristics of individuals which predict effective performance, or the relatively quick development of ability to give such a performance. It has already been stressed that the personality of the leader is not the only factor in the leadership equation, that the characteristics of the leader are important only in relation to the characteristics of the situation. But for any given leadership job or type of job, it is reasonable to suppose that the leader's personality is a major variable in effective leadership.

The earlier discussion of personal characteristics of group members which are relevant to group effectiveness points in a general way to the kinds of characteristics that might be significant in a leader in any situation: his more important needs, particularly as these relate to other people and to being a leader, and his characteristic ways of satisfying these needs or reacting to frustration; his attitudes, values and ideology concerning social relationships, power, authority, organizational goals, etc.; his knowledge and skills relative to technical problems, social relationships and self-understanding. Possibly other characteristics such as age, physique, bearing, social status, and so on, may be important too, since these may not only affect the leader's capacity to satisfy the demands of some situations, but more generally they may influence others' judgments of the probability that he can meet the situation effectively.

In order to further our understanding of the ways in which such characteristics may be related to leadership in particular situations, it would seem desirable to carry out intensive personality studies of leaders directly in conjunction with studies of their functioning in relation to other group members, rather than attempt to jump from personality studies on the one hand to criteria of leadership effectiveness on the other, as so many past studies have done. A beginning in the desired direction has been made by Jennings (**56**), and more recently, with more powerful personality instruments, by the First National Training Laboratory in Group Development held at Bethel, Maine (**27**).

Answers to a number of fundamental questions depend upon comparative investigations of this sort with reference to various types of group situations. To what extent, for example, do different leadership situations call for different personalities in leaders? Or, conversely, to what extent do leaders with given personality characteristics perform effectively in different situations? These questions are not necessarily answered by studies reporting different traits in leaders in different situations (see Stogdill's summary, **86**), the principal reason being that such traits or the measures of them may be determined to a great extent by the situation. Empirical studies of existing groups, involving more effective personality analyses and situational descriptions, would at least provide reasonable hypotheses as to the important relationships involved. These hypotheses might then be tested by rotating leaders of known characteristics through a series of different situations and determining the effectiveness of their performance. This is an undertaking, incidentally, for which a military organization provides unrivalled opportunities.

A related question having possible implications for leadership training might be investigated by a similar approach. To what extent in a given situation may leaders having different characteristics develop patterns of leader-follower interaction which are different in form but equal in overall effectiveness?

The problem of selecting leaders in a large organization must be thought of not only cross-sectionally, in terms of the different jobs to which an individual might be assigned at a given time and of the characteristics which he should accordingly possess at that time, but longitudinally as well, that is in terms of the individual's promoteability to positions of broader responsibility. Personality is rarely a fixed quantity, and so, in evaluating an individual's potentialities for filling a longitudinal series of leadership jobs (assuming they can be known in advance), it is probably unwise to demand of him all of the characteristics which they would require if considered cross-sectionally. But it is important to consider his potentialities for growth. In terms of research this points to the need for investigating leaders' careers, preferably through long-term, follow-up studies.

Methods of Studying Leader Personality

Most of the numerous studies of personality factors in leadership have been reviewed by Jenkins (**54**) and Stogdill (**86**). The record, as Jenkins remarks, is not a brilliant one. Part of the explanation for this, as previously suggested, may be found in an inadequate conception of leadership, in insufficient study of leadership behavior in specific situations, and in defective criteria. Another factor responsible may be the means employed to study personality. In the majority of cases leadership studies have used trait measures based either on pencil-and-paper tests or questionnaires, or on ratings, frequently ratings made by members of the leader's group. And the number of traits considered has frequently been very small. If, as seems probable, various characteristics or various personalities may make for effective leadership in a given situation, then consideration of but a few aspects of personality might be expected to yield meager results.

General developments in personality appraisal in recent years, and specific attention to the problems of officer selection during World War II, have made available a rich variety of techniques which, although in some cases not yet amenable to precise quantification, offer considerable promise for adequately comprehensive analyses of the personalities of leaders or potential leaders. Most of these techniques were embodied in the assessment program worked out by the Office of Strategic Services (**79**), in part on the basis of the procedures of the British War Officer Selection Boards (**36**). In addition to a number of standard pencil-and-paper tests and special questionnaires, the OSS program included personal interviews, projective tests, tests of physical daring and endurance, special tests of observation, memory, propaganda skills and other abilities thought necessary in OSS agents, sociometric tests and situational tests bearing on leadership and social relations. Of particular interest among those last mentioned are the leaderless group, assigned leadership, construction test (directing clumsy "stooges"), panel discussion, stress interview, dramatic improvisations (of roles assigned with reference to some fictitious situation), and an informal debate with hard liquor abundantly available. The tests were administered to candidates during a three-day stay at a country estate, during which the candidates had to maintain a fictitious identity, and during which, also, the tests were supplemented by a host of informal observations.

Conclusive evidence on the validity of the OSS procedures is lacking. Validity coefficients determined in various ways were relatively low. Partial responsibility for this, at least, was attributed to various factors other than the selection procedures themselves, notably the criteria. For the British program, and the similar program followed by the Australian Army, Gibb (**37**) has reported significant superiority in the performance of selected men at Officer Training Units. Data on combat performance are presumably not available.

From various other sources comes some additional evidence concerning certain of the specific procedures mentioned in connection with the OSS program. Both the Army and the Marine Corps have evidently found sociometric measures, based on nominations for leadership positions by fellow-members of a group, to be among the most effective single predictors of leadership success of the measures they have investigated. In the Marine Corps study (**95**), nominations by fellow officer candidates correlated as high as 0.47 with ratings given later in combat by superior officers. For West Point cadets, a correlation of 0.51 was found between similar sociometric measures and Officer Efficiency Report standings 18 months later (**5**). Of interest also is an experiment by the British Army in basing officer candidate selection on nominations by enlisted men, NCO's, and junior officers (**89**). This was reported to yield many suitable candidates who otherwise would have been neglected. Just what sociometric indices mean in terms of prevailing conceptions of personality is by no means clear, but they apparently do tap something fairly characteristic of individuals (**56**).

Other procedures which have received systematic attention are the interview and the biographical information blank. The Army, for example, found it possible to develop a quantified interview procedure involving no reference to other data, which had a reliability of 0.87 and a validity of 0.37 (against a nominating criterion), in contrast to a validity of about 0.12 for the usual Army board procedure in which all records about the candidate were available (**84**). A Biographical Information Blank also developed by the Adjutant General's Department had about the same validity (**82**). Chapple and Donald (**20**) have reported a method for recording and analyzing various characteristics of an individual's conversation in an interview, and have claimed considerable success in differentiating effective and ineffective leaders. Some evidence of the selection value of a stress interview procedure has been reported by Freeman *et al.* (**31**).

Finally we may note the studies of the personalities of successful business executives by Henry, Gardner, and their associates, using primarily the Thematic Apperception Test and short, undirected interviews. Although full data have not been published, preliminary reports (**44**) appear very promising.

These procedures seem to have in common a comprehensive approach to the individual personality, with an emphasis upon tendencies relevant to social participation. The evidence thus far available concerning their validity is not highly impressive, but the possibilities for development would seem to be. Further work on group situation tests, in particular, would be desirable. If these various techniques can be rationalized in terms of personality theory,[10] and evaluated with due regard to the significant aspects of particular leadership situations, the outlook for leader selection may not be too dark.

Leadership Training

Everyday experience, together with the results of a few small-scale experimental studies (e.g., **9, 66**), leave no doubt that in general the performance of leaders can be improved through training, but questions remain as to the best procedures for doing this and as to the limits of their effectiveness.

With respect to techniques, it appears more than probable that the inspirational lectures and readings sanctified by tradition, no matter how "streamlined" they may be in accordance with the latest public relations manual, are largely futile. One suspects—and the question could easily be answered—that the same holds true of the more recent trend toward "scientific" courses *about* human relationships or leadership. With leadership, as with technical training, it is probably practicing the job which counts. The outstanding development in this connection in recent studies of leadership training is the role-playing technique, whereby in small, informal groups individuals successively take

[10] It is not unlikely that personality theory will have to be re-rationalized in light of such studies.

the roles of leader and follower, with reference either to that group or to hypothetical situations, and under the critical eye of a participating leader-trainer and other group members (**9, 33, 34, 69, 97**). This is essentially an opportunity for practice, but it appears to differ in several important respects from practicing in a real leadership position or a conventional student leader position. The role-playing takes place in an atmosphere of make-believe, which makes possible greater freedom to experiment with new modes of behavior. Secondly, it involves filling, self-consciously and in rapid succession, the roles of both leader and follower. If leadership is basically a process of communication, and if effective communication rests upon anticipating the reaction of the other person (**23, 73**), then a procedure which encourages development of realistic expectations of the reactions of other group members should be productive. Thirdly, the procedure allows quicker knowledge of results than is ordinarily possible in real situations. This technique deserves intensive investigation, not only with reference to leadership training, but also because of its implications for a basic understanding of social interaction.

Other potentially fruitful kinds of experience suggested by some of the above studies are: discussion of the basic objectives and philosophy of leadership in the trainees' own groups, experiencing contrasting methods of leadership, observing movies of contrasting methods and of one's own performance, discussion of particular methods at appropriate intervals, and so on. In the application of all of these procedures, the skill of the leader-trainer seems to be a critical factor.

Studies of effective leadership behavior presumably determine the content of leadership training (and also the method). If, as was suggested earlier, different individual leaders can achieve equivalent results by different methods, and are predisposed to use such methods, training programs must probably allow for considerable flexibility. Investigations of the acquisition of particular leadership techniques, and of the effectiveness of different training procedures, as functions of trainee personality, would ultimately be desirable.

Leadership training cannot be, and in practice ordinarily is not, conceived simply in terms of special courses. Contacts with superiors and subordinates outside the classroom are likely to be more decisive than classroom experiences in shaping leaders' behavior (**41**). Thus, the improvement of leadership and leadership training calls for attention to on-the-job training procedures, and this may imply in turn the production of fundamental changes in organizational functioning through re-education of both leaders and followers. The possible roles of staff advisers in catalyzing such changes have been discussed recently by several writers (**51, 71**).

Summary

Because of the relatively undeveloped state of the science of group behavior, the attempt has been made in this chapter primarily to indicate areas and problems needing investigation—to raise questions rather than to answer them.

The term "morale" is assumed to refer to the effectiveness of group functioning in relation to group goals, or to aspects of group behavior indicative of the level of group effectiveness. Many different behavioral indications may be hypothesized as relevant in this connection, and various techniques exist or can be developed for observing these systematically. A prime necessity in research is the development of a precise and economical behavioral definition of group morale, through hypothesizing morale indicators, measuring these, analyzing the interrelationships of different measures, and confirming through further observation the utility of the factors or clusters thus discovered.

Given an adequate operational definition of morale, the way is open to investigation of

the variables related to it, and of ways of manipulating these variables in practical situations. Morale or group effectiveness may be conceived as dependent upon the group's characteristics—the needs, beliefs, skills, and so on, of individual members, and the patterns of interaction which obtain between individuals—and upon external circumstances affecting the group. Since these two classes of variables may not only indicate a present level of effectiveness (i.e., measure morale) but also influence a future level, their relationship to morale is largely a circular one. Among the manipulable variables of potential significance in military groups may be noted: the initial characteristics of members which are essential to any social participation, the patterning of personalities within the group, job assignment, training, length of time spent in the group, personnel changes, general conditions of group living, relations to other groups, special conditions of combat, group success and failure, and leadership. Because of the complex interrelationships involved, the experimental investigation of these variables encounters considerable difficulty in many instances.

Since the manipulation of (or allowance for) variables related to morale is in institutional groups primarily the responsibility of appointed leaders, the factor of leadership assumes central significance. A large share of research on this problem has been dominated by the conception of leadership as a personality trait or set of traits in the leader. The results have not been conspicuously successful. In the general conception now developing, leadership is viewed as a process of interaction between individuals, in which the characteristics of the leader or potential leader are important only in relation to the situation, that is the goals and other characteristics of the group and the relevant external circumstances. The urgent necessity in leadership research is for careful study of the leader-follower relationships which make for effective group performance, and of the differences in such relationships which characterize different group situations. Studies recently completed or now in progress suggest a variety of approaches to the analysis of these relationships.

Intensive analysis of leader-follower interaction is essential to intelligent research on the selection and training of leaders, since it lays a basis for adequate job analysis and classification, improvement of leadership criteria, and relevant hypotheses concerning leader personality factors and training procedures. Also of importance for research in leader selection is the exploitation of newer techniques for the overall study of personality, especially as it manifests itself in social participation. Studies of leadership training suggest the possibility of large improvements through use of appropriate training procedures. In large organizations ordinary extracurricular interaction is likely to impose critical limits on the effectiveness of special courses employing such procedures, a fact which calls attention to problems of on-the-job training and fundamental organizational change.

References

1. Allport, G. W. Morale and its measurement. *Publ. Policy, Harvard Univ.*, 1942, **3,** 3–17.
2. Anderson, H. H. Socially integrative behavior. *J. abnorm. soc. Psychol.*, 1946, **41,** 379–384.
3. Arensberg, C. M., & McGregor, D. Determination of morale in an industrial company. *Appl. Anthropol.*, 1942, **1,** 12–34.
4. Arrington, R. E. Time sampling in studies of social behavior: a critical review of techniques and results with research suggestions. *Psychol. Bull.*, 1943, **40,** 81–124.
5. Baier, D. E. Note on "A review of leadership studies with particular reference to military problems." *Psychol. Bull.*, 1947, **44,** 466–467.
6. Bales, R. F., & Gerbrands, H. The interaction recorder—an apparatus and check list for sequential content analysis of social interaction. *Hum. Relations*, 1948, **1,** 456–463.
7. Barnard, C. I. *Functions of the executive.* Cambridge: Harvard Univ. Press, 1938.
8. Barnard, C. I. *The nature of leadership.*

Cambridge: Harvard Univ. Press, 1940. Reprinted in: Hoslett, S. D. (Ed.) *Human Factors in Management,* Parkville, Mo.: Park College Press, 1946.

9. Bavelas, A. Morale and the training of leaders. In: Watson, G. (Ed.) *Civilian Morale,* Boston: Houghton Mifflin, 1942, 143–165.
10. Bavelas, A. Group decision. Paper read before Society for Psychological Study of Social Issues, 1943.
11. Benge, E. J., & Capell, D. F. Employee morale survey. *Modern Management,* 1947, **7,** 19–22.
12. Benne, K. D., & Sheats, P. Functional roles of group members. *J. soc. Issues,* 1948, **4,** 41–49.
13. Blankenship, A. B. Methods of measuring industrial morale. In: Hartman, G. W., & Newcomb, T. M. (Ed.) *Industrial conflict: a psychological interpretation.* New York: Cordon Co., 1939.
14. Bronfenbrenner, U. The measurement of sociometric status, structure and development. *Sociometry Monogr.,* 1945, No. 6.
15. Brown, J. F. A modification of the Rosenzweig Picture-Frustration Test to study hostile interracial attitudes. *J. Psychol.,* 1947, **24,** 247–272.
16. Burtt, H. E. *Principles of employment psychology.* (Rev. Ed.) New York: Harper, 1942.
17. Cattell, R. B. Concepts and methods in the measurement of group syntality. *Psychol. Rev.,* 1948, **55,** 48–63.
18. Chapin, F. S. *Experimental designs in sociological research.* New York: Harper, 1947.
19. Chapple, E. D. Personality differences as described by invariant properties of individuals in interaction. *Proc. nat. Acad. Sci.,* 1940, **26,** 10–16.
20. Chapple, E. D., & Donald, G. A method for evaluating supervisory personnel. *Harvard Bus. Rev.,* 1946, **24,** 197–214.
21. Child, I. L. Morale: a bibliographical review. *Psychol. Bull.,* 1941, **38,** 393–420.
22. Coffin, T. E. A three-component theory of leadership. *J. abnorm. soc. Psychol.,* 1944, **39,** 63–83.
23. Cottrell, L. S. The analysis of situational fields in social psychology. *Amer. sociol. Rev.,* 1942, **7,** 370–382.
24. Crannell, C. W., & Mollenkopf, W. G. Combat leadership. In: Wickert, F. (Ed.) *Psychological research on problems of redistribution,* Report No. 14 in AAF Aviat. Psychol. Program Res. Reports. Washington: Gov. Printing Off., 1947, 30–76.
25. Criswell, J. H. The measurement of group integration. *Sociometry,* 1947, **10,** 259–267.
26. Darley, J. G. War weariness and morale in air groups. In: Kelly, G. A. (Ed.) *New Methods in Applied Psychology,* College Park, Md.: Univ. Maryland, 1947, 129–133.
27. Deutsch, M., Pepitone, A., & Zander, A. Leadership in the small group. *J. soc. Issues,* 1948, **4,** 31–40.
28. Duff, I. F. *Medical study of the experiences of submariners as recorded in* 1471 *submarine patrol reports in World War II.* (Undated photostat copy prepared by Research Division, Bur. Med. Surg., Navy Dept.; classified as CONFIDENTIAL).
29. Estes, W. K., & Estes, K. W. Minnesota studies in war psychology: I. A set of miniature scales for the measurement of attitudes related to morale. *J. soc. Psychol.,* 1944, **20,** 265–276.
30. Franz, J. G. Survey of sociometric techniques, with an annotated bibliography. *Sociometry,* 1939, **2,** 76–92.
31. Freeman, G. L., Manson, G. E., Katzoff, E. T., & Pathman, J. H. The stress interview. *J. abnorm. soc. Psychol.,* 1942, **37,** 427–447.
32. French, J. R. P. Organized and unorganized groups under fear and frustration. *Iowa Studies in Child Welfare,* 1944, **20,** 229–308.
33. French, J. R. P. Retraining an autocratic leader. *J. abnorm. soc. Psychol.,* 1944, **39,** 224–237.
34. French, J. R. P. Role-playing as a method of training foremen. In: Moreno, J. L. (Ed.) *Group psychotherapy: a symposium.* New York: Beacon House, 1946, 172–187.
35. Freud, S. *Group psychology and the analysis of the ego.* London: International Psycho-Analytical Press, 1922.
36. Garforth, F. I. de la P. War Office Selection Boards (O.C.T.U.) *Occup. Psychol.,* 1945, **19,** 96–108.
37. Gibb, C. A. The principles and traits of leadership. *J. abnorm. soc. Psychol.,* 1947, **42,** 267–284.
38. Golden, C. S., & Ruttenburg, H. J. *The dynamics of industrial democracy.* New York: Harper, 1942.
39. Greenwood, E. *Experimental sociology: a study in method.* New York: King's Crown Press, 1945.
40. Guttman, L. A basis for scaling qualitative data. *Amer. sociol. Rev.,* 1944, **9,** 139–150.
41. Haire, M. Some problems of industrial training. *J. soc. Issues,* 1948, **4,** 41–47.

42. Harding, J. The measurement of civilian morale. In: Cantril, H., *et al. Gauging public opinion.* Princton: Princeton Univ. Press, 1944, 233–285.

43. Hemphill, J. K. *Situational factors in leadership.* Ph.D. Thesis, Univ. Maryland, 1948.

44. Henry, W. E. The business executive: the psychodynamics of a social role. *Amer. J. Sociol.*, 1949, **34**, 286–291.

45. Information and Education Division, Army Service Forces. What is morale? *What the soldier thinks*, 1943, No. 1. 1–7.

46. Information and Education Division, Army Service Forces. Attitudes toward noncommissioned officers. *What the soldier thinks*, 1943, Quarterly report No. 2, 41–45.

47. Information and Education Division, Army Service Forces. Job satisfaction and the utilization of skills. *What the soldier thinks*, 1943, Quarterly report No. 2, 47–53.

48. Information and Education Division, Army Service Forces. How WAC's view their officers. *What the soldier thinks*, 1945, No. 15, 7–9.

49. Information and Education Division, Army Service Forces. "Square deal" or not? *What the soldier thinks*, 1945, No. 16, 1–5.

50. Information and Education Division, Army Service Forces. Things combat soldiers would change in the Army. *What the soldier thinks*, 1945, No. 16, 12.

51. Jaques, E. Social therapy: technocracy or collaboration? *J. soc. Issues*, 1947, **3**, 59–66.

52. Jenkins, D. H. Feedback and group self-evaluation. *J. soc. Issues*, 1948, **4**, 50–60.

53. Jenkins, J. G. The nominating technique: it uses and limitations. Paper presented before Eastern Psychological Association, Atlantic City, 1947.

54. Jenkins, W. O. A review of leadership studies with particular reference to military problems. *Psychol. Bull.*, 1947, **44**, 54–79.

55. Jennings, H. H. Control study of sociometric assignment. *Sociometric Rev.*, 1936, **1**, 54–57.

56. Jennings, H. H. *Leadership and isolation; a study of personality in inter-personal relations.* New York: Longmans Green, 1943.

57. Kardiner, A., & Spiegel, H. *War stress and neurotic illness.* New York: Hoeber, 1947.

58. Kerstetter, L. M., & Sargent, J. Reassignment therapy in the classroom as a preventive measure in juvenile delinquency. *Sociometry*, 1940, **3**, 293–306.

59. Knickerbocker, I. Leadership: a conception and some implications. *J. soc. Issues*, 1948, **4**, 23–40.

60. Kornhauser, A. Psychological studies of employee attitudes. *J. consult. Psychol.*, 1944, **8**, 127–143.

61. Krech, D., & Crutchfield, R. S. *Theory and problems of social psychology.* New York: McGraw Hill, 1948.

62. Leavitt, H. J., & Adler, N. *Validation of officer selection tests by means of combat proficiency ratings.* Medical Field Research Laboratory, Camp Lejeune, N. C., M. and S. Res. Prog. No. X-620 (Sub. No. 135) 16 May 1946.

63. Lewin, K. Time perspective and morale. In: Watson, C. (Ed.) *Civilian morale.* Boston: Houghton Mifflin, 1942, 48–70.

64. Lewin, K. Group decision and social change. In: Newcomb, T. M., Hartley, E. L., *et al.* (Eds.) *Readings in social psychology.* New York: Holt, 1947, 330–344.

65. Lewin, K. Frontiers in group dynamics. *Hum. Relations*, 1947, **1**, 5–41, 143-153.

66. Lewin, K. Lippitt, R., & White, R. K. Patterns of aggressive behavior in experimentally created social climates. *J. soc. Psychol.*, 1939, **10**, 271–299.

67. Lippitt, R. An experimental study of the effect of democratic and authoritarian group atmospheres. *Iowa Studies in Child Welfare*, 1940, **16**, 43–195.

68. Lippitt, R. The morale of youth groups. In: Watson, G. *Civilian morale.* Boston: Houghton Mifflin, 1942, 119–142.

69. Lippitt, R. The psychodrama in leadership training. *Sociometry*, 1943, **6**, 286–292.

70. McGregor, D. Conditions of effective leadership in the industrial organization. *J. consult. Psychol.*, 1944, **8**, 55–63.

71. McGregor, D. The staff function in human relations. *J. soc. Issues*, 1948, **4**, 5–22.

72. McNemar, Q. Opinion-attitude methodology. *Psychol. Bull.*, 1946, **43**, 289–374.

73. Mead, G. H. *Mind, self and society.* (Ed. by C. W. Morris) Chicago: Univ. Chicago Press, 1934.

74. Miller, D. C. The measurement of national morale. *Amer. sociol. Rev.*, 1941, **6**, 487–499.

75. Moreno, J. L. *Who shall survive?* Washington: Nervous and Mental Disease Publishing Co., 1934.

76. Moreno, J. L., & Jennings, H. Advances in sociometric technique. *Sociometric Rev.*, 1936, **1**, 26–40.

77. Moreno, J. L. Sociodrama, a method for the analysis of social conflict. *Psychodrama Monogr.*, 1944, No. 1.

78. Newcomb, T. M. News and morale: a miniature experiment. In: Watson, G. (Ed.)

Civilian morale. Houghton Mifflin, Boston, 1942, 175–185.

79. O.S.S. Assessment Staff. *Assessment of men.* New York: Rinehart, 1948.
80. Pigors, P. *Leadership or domination.* Boston: Houghton Mifflin, 1935.
81. Proshansky, H. M. A projective method for the study of attitudes. *J. abnorm. soc. Psychol.*, 1943, **38**, 393–395.
82. Richardson, M. W. Selection of Army Officers. In: Kelly, G. A. (Ed.) *New methods in applied psychology.* College Park, Md.: Univ. Maryland, 1947, 79–85.
83. Roethlisberger, F. J., & Dickson, W. J. *Management and the worker.* Cambridge: Harvard Univ. Press, 1939.
84. Rundquist, E. Q. Development of an interview for selection purposes. In: Kelly, G. A. (Ed.) *New methods in applied psychology.* College Park, Md.: Univ. Maryland, 1947, 85–95.
85. Staff, Personnel Research Section, the Adjutant General's Office. *The forced choice technique and rating scales.* (Paper read before Amer. Psychol. Assn., 1946).
86. Stogdill, R. M. Personal factors associated with leadership: a survey of the literature. *J. Psychol.*, 1948, **25**, 35–71.
87. Stogdill, R. M., & Shartle, C. L. Methods for determining patterns of leadership behavior in relation to organization structure and objectives. *J. appl. Psychol.*, 1948, **32**, 286-291.
88. Sullivan, H. S. *Conceptions of modern psychiatry.* Washington, D. C.: The William Alanson White Psychiatric Foundation, 1947.
89. Sutherland, J. D., & Fitzpatrick, G. A. Some approaches to group problems in the British Army. In: Moreno, J. L. (Ed.) *Group psychotherapy: a symposium.* New York: Beacon House, 1946.
90. Terman, L. M., & Buttenweiser, P. Personality factors in marital compatibility. *J. soc. Psychol.*, 1935, **6**, 143–171, 267–289.
91. Van Dusen, A. C. Measuring leadership ability. *Personnel Psychol.*, 1948, **1**, 67–79.
92. Vaughn, C. L. The nominating technique. In: Kelly, G. A. (Ed.) *New methods in applied psychology.* College Park, Md.: Univ. Maryland, 1947, 22–26.
93. Ward, L. B. Criteria of crew proficiency in operational training. In: Kelly, G. A. (Ed.) *New methods in applied psychology.* College Park, Md.: Univ. Maryland, 1947, 37–42.
94. Whyte, W. F. *Street corner society.* Chicago: Univ. Chicago Press, 1937.
95. Williams, S. B., & Leavitt, H. J. Group opinion as a predictor of military leadership. *J. consult. Psychol.*, 1947, **11**, 283–291.
96. Witherspoon, M. M. Morale in the new world war. *Proc. U. S. Nav. Inst.*, 1940, **66**, 675–684.
97. Zander, A. F. Role playing: a technique for training the necessarily dominating leader. *Sociatry*, 1947, **1**, 225–235.
98. Zeleny, L. D. Sociometry of morale. *Amer. sociol. Rev.*, 1939, **4**, 799–808.
99. Zeleny, L. D. Selection of compatible flying partners. *Amer. J. Sociol.*, 1947, **52**, 424–431.

PART VIII.

SELECTION AND TRAINING

Chapter 23

SELECTION AND TRAINING

LEE J. CRONBACH

University of Illinois

AND

W. D. NEFF

University of Chicago

Introduction

It is unusually difficult to summarize in a short space the present status and prospective problems in selection and training. So much is now known and so much remains to be known that volumes can be written about purely military problems, disregarding more general principles. The most that can be attempted, therefore, is a strategic overview of the topics and problems that make continued research in these areas vital.

A second difficulty is that of separating fundamental theory from administrative problems. What research should be done on selection depends upon how undersea warfare will be organized—for example, will men be screened and classified before any training, will officers generally come from the regular Navy, will there be periodic checks on shipboard for reclassification of men? The procedures to be developed must always be designed with relation to some practical situation. For that reason, some of this chapter discusses practical plans for selection and training, rather than scientific theory alone.

There are few problems of selection and training in undersea warfare not equally prominent in all other branches of the service. How one tests men, how one trains men, how one assesses misfits, and so on, will be much the same, whether one is choosing fire-control personnel for a shore battery, officers for a tactical staff, or toolmakers for an ordnance works. The particular tests or training materials will vary, but the psychological principles remain the same. Our present theory of military selection and training draws very heavily on work in public schools, colleges, and industry. It is essential to have a coordinated attack upon basic problems, conceived to answer each question in the broadest possible terms. Research specific to aviation selection, submarine selection, or rifleman selection can be attacked more readily after the problems common to all these areas have been solved.

Selection

Problems of Mass Processing

Handling manpower on the scale required by military processing is a problem nowhere encountered in peacetime. Yet the broad planning of a classification program has a logic of its own that can be arrived at by scientific processes.

Principles of Placement

Scientists evaluate a test by its "validity coefficient", a number showing how well the test score compares with the man's later work on the job. On the basis of early experience and from mathematical reasoning, psychologists set the rule: "A test with a validity coefficient of .50 or .60 is good enough for identifying good and poor men." By this rule, one could decide whether any proposed aptitude test should be used in selection. But a reconsideration of the problem by Taylor and Russell (**45**) before World War II suggests a more realistic standard. As they point out, it is obvious that a test with any validity at all is more useful for classification than a blind guess, or taking every tenth man out of line to fill a job. They derived tables to show just how much the quality of a working force would be raised if a test of given validity is used. This led to the conclusion that tests of very low validity can be helpful and profitable if it is possible to skim off the cream of the recruits. If only one recruit out of 100 is needed for a draft for School A, a test with validity .10 will raise the quality of men sent to the school; if half the recruits must be sent to School A, a test with validity of .10 will not help enough to be worth using.

In selecting, one may choose 10 men out of 100 for a job and discard the rest. In placement, including military training, one must place 99 men out of 100 in some job or other; very few men can be discarded. Placement is a different problem from selection. It no doubt has its mathematical logic which, if formulated, would provide a guide for choosing or rejecting tests. This logic must be worked out.

In broad terms, the problem is this: One has x rates (or specialties, or stations) to fill; some of them are more important tactically than others, and error in filling them must be avoided. For each rate, selection methods are available which predict the success of each man in each job, with a certain margin of error. What processing formula can consider both the prediction and the importance of the job, to place each man so that the total force will have maximum effectiveness? If all men not better fitted for something else must be assigned to the job lowest in importance, it does not appear worthwhile to test aptitude for that job even if a test of high validity is available. For the most important job, a selection method should be used even if its validity is so low that it would decrease failures by only one percent. What is the relation between test validity, job importance, proportion of recruits required for each job, and other variables? This is a mathematical question, open to research, which would suggest how to design processing routines in wartime (**16,** pp. 58–61). It might lead to definite rules, e.g., "One should fill the quota for the most important job with the best men before making any other assignment," or "The classifying office should place every man in the specialty for which he is best fitted until half the billets in each specialty are filled; after that, quotas should be filled in order of importance."

All present statistical methods predict a man's success in one field at a time. Problems of differential prediction likewise require statistical solution. The classifying officer must consider the man's relative success in many fields. Suppose a man has an aptitude score of 60 for torpedoman, and an aptitude score of 60 for radioman. He is not likely to do equally well in both jobs; errors of measurement and errors of prediction make him a better risk for one job or the other. No present statistical routine is adequate for giving a score showing "likelihood that this man will be a better torpedoman than radioman." Yet that is predictable, and it is a judgment crucial in manpower utilization. Probably it will use different tests from the present single-job prediction; intelligence is needed for both jobs, but an intelligence test would not help decide to *which* job to assign the man. The only direct helps now available on the differential prediction problem are a set of formulas on the reliability of test profiles, recently simplified by Bennett and Doppelt into practical form (**4**), and a preliminary proposal by Brogden (**8**) for the determination of differential cutting scores. A thorough mathematical study of prediction is called for.

Billet Analysis

Processing fills billets, but the psychological problems begin earlier, when billets are defined (**60**). What constitutes a job?

Ideally, a job would be a set of activities calling for common aptitudes and knowledge, so that the man who is good at one aspect will be good at all aspects of the job. The activities for any single billet must be so arranged in time and space that one man can perform them all, and the total organization of billets for the ship must cover all duties to be done. An efficient billet design makes use of one man's skill as broadly as possible. It was very inefficient to give the same advanced training in electronics to sonar materiel men, radar technicians, and radio technicians. The recent change, establishing the new rate of Electronic Technician's Mate, makes it possible to train one man to perform duties three men formerly had to learn. But billets are rarely organized on a psychological basis. In a submarine, the radioman normally becomes a sonarman when the boat submerges. The reason? Because the sound gear is accessible to him, and the radio cannot operate under water. But radio aptitude may not make a good sonarman. A sonarman must have tactical visualization aptitudes, keen hearing for pitch and for high tones, and other abilities the radioman does not use. Perhaps on some ships a signalman or radarman would get more information out of the sound stack than a radioman, if he were properly trained. Since it is not possible to fill every billet with a superman who can do everything perfectly, it is important to design billets so that the task has psychological unity, calling for similar aptitudes and personal characteristics.

Selection with Several Tests

The traditional method of making predictions is as follows: The test is sought which predicts job success best, then another test of a different function is sought which, added to the first, gives more exact prediction. This process of adding tests to make an accurate composite prediction is called multiple regression. Although this is a well-understood method, it has a serious limitation which under some circumstances leads to unsound processing. In ASW, early in the war, this procedure led to a formula in which pitch judgment and mechanical aptitude were averaged, to estimate probable success in sonar school. Many good men were sent to sonar school, where they failed; they had been passed when very high mechanical insight counterbalanced their very poor pitch judgment. Such men would have succeeded in radar or radio school, but could not make sonar judgments. The multiple-regression method assumes that strength in one ability compensates for weakness in another; this is often untrue.

The solution which avoids the assumption of compensation is the multiple cutoff. Men with poor pitch are discarded; men with poor mechanical comprehension are discarded; further cutoffs are made with other tests until only men without deficiencies remain. The cutoff method has not been studied very thoroughly. It has limitations of many sorts. Thorndike has explained why the Air Force chose to base pilot selection on the multiple-regression method (**46**, pp. 89–91), but has not considered the many situations where it leads to poor selection. Ruch (**40**) has found definite evidence for the superiority of the cutoff method under some circumstances. A careful study is needed to define the precise conditions under which it is advisable to use multiple cutoffs rather than a composite formula for prediction. Work such as that of Springbett, who devised an efficient method for determining multiple cutting scores (**43**), needs further development.

Shipboard Classification

A major practical problem in World War II was shipboard classification. Processing to provide the fleet with excellent men in all rates was carried out in training centers. As a supplement to this system, shipboard strikers were often permitted to try for a rate even when they lacked necessary aptitudes. Probably many men were rated at sea whose skill on the job was inferior. Non-rated assignments were made with very little consideration of ability. What is required is a set of valid classification data which can follow the man to each station and can be used there for making further sssignments. It is essential that these data be used, and that periodic checks be made to assure that men classified on shipboard are competent in their rates. Standardized examinations for advancement in rating can be developed (**44**). More work of this type is needed in order to upgrade the quality of shipboard strikers. Development of this program is an administrative problem.

Utilization of Personnel

A corollary weakness to the placement of inadequate men in important posts is the wastage of trained men. In World War II, some of this was unavoidable due to strategic developments. But many times men were graduated from a school and sent to duty where their newly acquired skill was not called for. The Army developed a personnel audit procedure to make sure that men were not retained on jobs which less trained men could do. While these methods met a wartime problem, it is probable that interim research can develop a more effective plan for utilization of superior men.

Judgment in Classification

A serious problem for the armed services is whether classification specialists should be allowed to use judgment in disposing of cases. One could average a man's scores and assign him without thought, or one could consider supplementary facts about him in making an assignment. All logic and psychology would be on the side of the second alternative, since use of one's intelligence should improve prediction. But the evidence from two wartime studies (**10, 16,** pp. 287–288) shows that, under service conditions, whenever classifiers departed from the disposal warranted by test scores, they made poorer selections than if they had followed the scores blindly. This should not be; research is required to determine exactly why "intelligent" departure from test results goes wrong, and how judgment can be used to improve rather than decrease efficiency of personnel placement.

Tests of Aptitude

In every characteristic we study, men vary greatly. This is not just a matter of their training or their willingness to do well; performance varies because of differences in aptitude which can be overcome, if at all, only by long and arduous training. It is common for the best men in any aptitude to be three to six times as good as the poorest men. Thus, the best lookouts detect targets at four times the range of the poorest lookouts (**39**). Men near the average are not much different from each other, but men at the extremes are quite exceptional. Suppose, on a typical test, two-thirds of the men fall within 10 points of average. Then 15 percent more will be from 10 to 20 points above average, and there is a small group of outstanding men who are as much as 30 points beyond average. If these men are entrusted with the most difficult tasks, they will perform them much better than the average or slightly superior men. Because there is so much difference between the top five percent and the men who merely reach the top quarter of a group, it is essential that aptitude testing identify the truly best men. Fortunately, the man best in one specialty is not likely to be best in another. Most technical jobs require intelligence, but in addition different jobs make special demands for coordination, sense of rhythm, vision, mechanical knowledge, and so on. There is a great difference in effectiveness between a force where there is a good man on every job and a force where the man on each job is the very best man for that duty. It is the latter force whose lookouts sight airplanes at extreme range, and whose pilots make carrier landings in a rough sea with minimum accidents.

Aptitude testing has been of great service in upgrading military forces. Air Force experience in pilot selection is an example: in a group of men screened only by a physical examination, the failure rate was 77 percent, but among men selected by the best combination of tests, the failure rate was only 24 percent (**16,** pp. 75 ff.). Even this is far short of perfect prediction; the validity coefficients around .65 leave much room for improvement.

Identification of Basic Aptitudes

One major line of progress has been the factorial technique for analyzing tests. In the past, testing has been a trial-and-error process. One test after another is tried to measure various components of job success, and those which work are retained. This leads to several types of inefficiency, well represented in Peterson's study of the Navy Classification Battery (**57,** pp. 19–21). (1) Tests supposed to measure different abilities really measure the same thing under different names. This was true for GCT and Reading. (2) Classification personnel are directed to consider scores on separate parts of a test, when the separate parts have no real meaning (for example, Mechanical Knowledge: Mechanical, and Mechanical Knowledge: Electrical were nearly identical psychologically). (3) A test supposed to measure one factor may measure something else, so that the intended factor is left unmeasured. Thus, the Arithmetic test turned out to be a measure of general schooling and intelligence, and no measure of quantitative ability was in the basic battery.

Factor analysis is more than a technique for removing inefficient tests from a battery, though that outcome is significant when it permits a classification interviewer to consider only three test scores per man instead of twelve. The greater contribution of factor analysis is that it maps out the relations between abilities so that future research can be more systematic. If we know what factors (common elements) are found in tests which seem to work in job prediction, we can design tests to measure those factors directly. Some investigators believe that tests could be made for all the significant abilities of men, and that in classification for any given job one would merely combine such scores to measure the combination of abilities needed on that job (**19**).

Studies such as that of the Air Force, which identified nearly 30 factors in their classification tests, are to be encouraged, since this yields a listing of abilities which ought to be considered in setting up any future classification plan. It would be most helpful in planning selection of men to operate any new gear to have an efficient list of all major human abilities which might be involved. But study of aviation tests is not enough. Studies of all types of tests found useful in classification are needed. Probably a good place to begin such a study would be to assemble all tests now known to be of use in processing, whether from Navy, Air Force, industry, or schools, to administer overlapping sets of them to similar groups of men, and to separate out the factors.

One of the fundamental issues remaining to be settled in factor analysis is: what mathematical system should be employed? There are a large number of mathematically sound procedures, of which mention may be made of Thurstone's centroid method leading to correlated factors, Guilford's use of the centroid method to get independent but less meaningful factors, Holzinger's extraction of a "general" factor before isolating specific aptitudes, and Kelley's principal components technique. Which of these is now used depends on an investigator's preference. Probably, for military purposes, there is one method which is most efficient. Since factorial studies made in the next twenty years will no doubt affect all future processing, it is especially desirable to have some statistician analyze the military problem to determine which plan should be followed in current studies. All methods are sound, but almost certainly one of them will be of greatest use in designing selection batteries.

Factor analysis is not a path to be followed blindly. It can identify abilities found in present tests, but will overlook abilities not now tested. An unending search for further abilities must be made; for example, a psychologist could profitably employ his time inquiring "what types of individual differences in hearing capacities might conceivably be tested?" Work prior to World War II had stemmed from Seashore's interest in music. He had identified numerous hearing aptitudes found in music, and had made tests. Two of them, the rhythm test and the pitch test, turned out to be useful in selecting code operators and sonar operators, respectively. These tests were used because someone had designed them; no one knows how many other significant hearing aptitudes were overlooked because group tests had not been considered worth developing. Searches for types of differences, without regard to their present practical significance, may well be made in motor ability, reasoning, vision, and so on.

Preliminary research is needed to demonstrate the effectiveness of factorial test batteries. While many workers have hopes that combining "pure" tests of different factors will give effective prediction, this has not been tried. It may be that such a battery will be inefficient in classification, if it requires some men to take dozens of tests to get the same information that one test designed to predict a single job could do. It is imperative that workers educated in factorial methods set up a small-scale classification program to determine what theoretical and practical problems will arise. Such a program must be carefully validated. Factor studies so far have shown that factorial tests measure different things, but no study has yet shown that the factor scores give superior predictions of anything.

A departure to be encouraged is factorial studies based on criteria of job performance. The present technique is to make a map, or coordinate system, which most efficiently describes the tests used in selection; then a given job is located relative to the test factors. It might be more efficient, although more difficult, to design a factor pattern to describe Navy jobs most efficiently, and to locate tests in terms of job factors. This is so radical a departure that no methodology exists. It might be profitable enough to justify even the most extreme labor. So far, no study has considered more than one criterion at a time. If all Navy jobs were considered at once, and a listing were made of the factors most prominent in all Navy jobs, an exceedingly useful classification battery could be designed. The effect of this plan on test construction is discussed below.

Factorial Design vs. Job Replicas

One of the gravest doubts about the factorial approach is that it tends to disregard "unique" abilities, found in only one test. The Air Force list of factors reports elements found in several tests, but it discards, for example, the special factor found in the complex coordination test. This test, in which the pilot operates a stick and rudder in response to light signals, had nearly the highest validity of any test in the program (**18**). But 27 percent of the saturation of this test was a factor found only in this test, a factor which does not appear in the list resulting from analysis. Other tests having superior validity, but heavily influenced by factors not in the master list, are Map Distance (for navigators), Biographical Data (pilot), and Discrimination Reaction Time (valid for pilot, bombardier, and navigator). Tests known as job replicas, which are complex tests closely resembling the job to be predicted, have frequently been excellent predictors. Factor analysts assume that tests of elements, rather than of complex combinations of elements, are the most efficient tests. But when one integrates many abilities into a complex pattern, he does more than add them. Concentrated research is necessary to determine whether factorial tests can reproduce the predictive effectiveness of job replicas. If they cannot, research should determine for which jobs replica tests should be designed, and should if possible establish general principles for designing such tests. Job replicas are costly and time-consuming; they should be used only where indispensable. But, if, for any Navy job, they improve the probability of supplying good men to the Fleet, they are indispensable.

One type of job replica, the learning sample, was found to be a good predictor during the war for code operators and gun trackers (**5, 33, 59**). Men who learn slowly in early trials are poor prospects. Research is needed to compare the selective efficiency of learning samples and conventional aptitude tests. If learning samples are superior, it may be necessary to redesign classification procedures to allow dropping and reassignment of men after they have started a school course.

Efficient Test Design

The design of psychological tests has rarely been studied systematically. A few significant principles have emerged, but it is probable that a search for general laws of test design would turn up many more. As examples of the sort of knowledge to be sought may be mentioned the current theory that time limits should be long enough that most men will finish all they can do, the development of efficient formulas for item analysis as a technique for improving tests (**9**), the identification of item forms which are subject to response sets and should be avoided (**12**), and the study of the relation between item difficulty and test reliability (**21**). The last named, for example, shows that a test to screen out the poorest 20 men per 100 should be composed of items which just 80 percent of the men can pass, rather than a mixture of easy and hard items. It is rules such as these that permit testers, measuring a new variable, to construct a test that accurately measures men in a short time.

It would be profitable to extend studies of test design in several directions. One approach is the development of tests which do not depend on reading. The Bureau of Naval Personnel made effective use of tests in which men assembled actual gear, but such tests are hard to administer. Perhaps a better solution is pictorial tests (**17**). In one extremely provocative study at Great Lakes Training Center, a picture test for gunners and a traditional verbal achievement test were given. The verbal test turned out to be a measure of reading ability; the picture test was a far better measure of ability to operate a gun (**54**). Research should determine the conditions under which picture tests are needed, how they should be designed, and what limitations they have.

Group tests are important to the armed services. Nearly every individual test would be more useful if it could be given accurately to groups of men. Research may be able to develop a guide for converting individual tests into group tests. In the absence of this, it is imperative that an attempt be made to develop a group form of every promising individual test. Examples of such adaptations are the group audiometer test and the projection eikonometer (**23, 57,** pp. 77–79).

Another type of research would stress factor analysis of single tests, item by item. The usual method of improving a test discards items which differ from the remainder of the test. In industrial refining, sometimes by-products removed from a substance turn out to be more important than the main material being purified. In tests, likewise, study of the rejected items might call attention to valuable test elements now being ignored.

Instead of improving tests merely by internal analysis, it is possible to choose test items so that each specific item is as good a predictor as possible (**20**). This in itself should improve tests for single selection tasks (e.g., CIC aptitude). But if studies of job criteria show certain abilities (say, mechanical comprehension, arithmetic computation, etc.) to be important in many jobs, it would be more desirable to make one test of each ability than to have the factor tested over and over for each job. Yet the specific type of arithmetic important may vary from job to job. One task calls for skill with fractions, one for decimals; one uses little division, one uses much. By Gulliksen's method (**20**) one could establish separate keys for the same test, each key giving the estimate of arithmetic ability that correlated best with one single job. Research can determine how much predictive validity is increased by establishing multiple empirical keys for a single standard aptitude test.

In this connection, an important lesson arises from a BuMed study. Aviation psychologists concerned with predicting pilot success had found a biographical inventory to be a helpful selection tool. This is in accord with many other studies, but it should be noted that life histories are useful only after specific research has shown this or that item to be correlated with success. Permitting a classification officer to assign men merely on the basis of his opinion as to what past background is desirable produces assignments based more on superstition than on science. In the BuMed study, the investigators followed up a guess that some items in a biography were effective predictors for one group of men but not for others. Their investigation showed that there were 70 items in the inventory which identified, among men of high ability, those who would pass the training and those who would fail. There were 85 items which identified prospective failures among those with low ability test scores. But only seven items were predictive at both ability levels. As a result, the psychologists developed three separate keys for the Biographical Inventory, to be used at different ability levels (**55**). Applying to a man's inventory the key for his ability level raised the validity coefficient only by a few points, it is true; but even such gains cumulate to major savings in military processing. The procedure in this study seems worthy of further trial with other predictors. This multiple-key technique should certainly be applied to empirical personality tests.

On this and other problems in military processing investigators will find the monograph, *The Prediction of Personal Adjustment* (**28**), informative and provocative. Horst and his collaborators listed, in 1941, many crucial unsolved problems in prediction, and suggested fundamental studies needed. They frequently developed tentative mathematical solutions to such problems as item weighting. Their suggestions were rarely developed adequately for application in World War II, but their monograph is an excellent point of departure for many of the studies proposed in the present report.

Psychomotor Functions

It is very important in military processing to measure motor aptitudes: coordination, speed of movement, steadiness, etc. Tests of this sort have often been useful, especially in selecting aviation personnel, but they are expensive to design and administer. Unfortunately, different jobs call for different motor abilities; no two or three motor tests can measure all the factors essential in key jobs. It would be desirable to have systematic research on the testing of psychomotor functions to determine ways in which they may be used effectively in mass processing. One possibility is to devise pencil-paper tests which will test motor abilities. For functions not so treatable, it may be possible to devise an apparatus which will quickly and efficiently give many motor aptitude scores. In the past, motor tests have been designed individually, to test some one ability. Probably design could be improved so much that many men could be tested at once by a single examiner, and a single testing could yield all the needed aptitude scores.

Training to Overcome Low Aptitude

Aptitude testers are inclined to think of abilities as fixed and not subject to training. Actually, many of the deficiencies men show on aptitude tests can be remedied by suitable training. Sometimes this training is quite simple; some people improve in motor coordination just by becoming familiar with the test apparatus (**50**). Usually defects are difficult to remove, but appropriate training can improve them. Thus, Wyatt demonstrated that pitch judgment can be trained, even in adults who seem to have poor discrimination (**51**). It would be very valuable to know how much training is required to overcome deficiency in each important aptitude. Then, in a given situation, one could decide whether it is more efficient to screen deficient men and send them to other duty, or to admit them and remove the deficiency by special work. The Fleet Sonar School at Key West, during the war, required all men with deficiencies in Doppler (pitch judgment) tests to attend special drill sessions in addition to the regular program. As a result, men were salvaged who would otherwise not have qualified. It may be more practical to improve mechanical comprehension, or visualization of relative movement, or even ability to see in dim light, than to reject generally competent men. Only research on the ease or difficulty of remedial training can answer such a question.

Personality Assessment

Aptitude tests determine what a man can do at his maximum; a much greater and more significant problem is to determine what he is commonly likely to do. This is the problem of identifying his habitual ways of responding, which we can call his personality. Personality tests are most familiar as tools to identify misfits and potential troublemakers, or men who will cave in under stress. This is only a small fragment of the problem which must be studied. Personality tests, if they were good enough, should tell which men will be most alert on routine patrol watches, which men will show greatest self-possession in a crisis, which men are so "perfectionist" that they will keep their equipment in perfect order, and so on. Any research which studies how men differ in the way they do their jobs should help us make use of a factor of which we are now ignorant.

Previous research on personality in military operations has been very largely confined to two topics: the separation of emotionally healthy men from potential breakdown cases, and the identification of potential leaders. While these are important and difficult lines of research, both are based on the too simple conception that men with "good" personalities are much alike, so that after such superior men are identified, research on personality is finished. That this is not so, and that men of the high quality obtained by the submarine service are extremely variable in personality, is demonstrated dramatically by a remarkable document of the past war (**14**). USS TANG, at the end of a patrol, fired its twenty-fourth and last torpedo at a target. The torpedo made an erratic run, turned, and exploded against the after portion of the submarine. The submarine went down by the stern, trapping the crew at about a 50-foot depth. In the confusion of the crash, in the subsequent assembling of the men in the forward compartment, and in the execution of escape operations, the behavior of men under stress revealed wide individual differences. The survivors' accounts are a dramatic and internally consistent narrative of rare psychological value. One incident deals with two men luckily trapped with their heads in an air bubble beneath the superstructure. They planned to make a free ascent through an open hatch, the enlisted man clinging to the lieutenant's legs, since only the latter knew where to find the opening. Yet as the officer swam out, the man loosed his grip and was never seen again. What difference in personality causes one man to carry out a plan in an emergency, and causes another to give up?

In the forward compartment, reports are more complete. In one scene, four men were in the escape bell, ready to come through the port and ascend to the surface on "lungs."

But the tank of oxygen was exhausted, and only one man inflated his apparatus with air and made the ascent. The others, unwilling to make the ascent on air, even though it is known to be practicable, remained in the bell, were taken back into the sub, and two of the three made no further attempts even though they were certain to die in the sunken boat. The third man directed solution of the oxygen problem and led the next escape. In one group, two officers were unable to escape, one due to exhaustion and the other due to both exhaustion and a minor injury. The former remained in the ship and told the men they could try to escape if they wished, but he was through trying. The other rested and escaped in a later group. Still another case is that of the officer who, despite wounds about the head, was self-possessed enough to direct destruction of confidential papers and later to make the ascent, but loaded upon himself two pistols and a bayonet, presumably to do battle with the destroyers topside.

This revealing tragedy draws attention to the great differences that lie beneath the surface of men who are an excellent group of brave and well-qualified submariners. The more ordinary crises of undersea warfare rarely reveal personality contrasts so sharply, but it is probably true that under depth-charging, crash dives, battle surface, or other stress conditions, some men are more disorganized than others. This is a most difficult task order for prediction; not that we need to predict which men will escape under the rare conditions the *Tang* experienced, but because we need to know which men keep their heads in emergencies or even outdo themselves, which men perform without slackening even when the odds are heavily against them, which men take the hazardous course rather than the cautious one, etc.

Screening Devices

The first and greatest use of personality tests is to screen out undesirable men. Such tests (or, rather, questionnaires) have been widely used for screening of recruits, along with psychiatric interviews. Any number of studies have proved that such tests do identify most of the recruits who should be interviewed with care (**57**, pp. 36–52). But such tests are much less effective in dealing with superior groups, such as submariners (**3**). This does not show that personality is irrelevant; it only indicates that more powerful tools are needed to identify excellent prospects than to screen out poor ones.

There are four basic approaches worth consideration as ways of bettering personality screening. One is the forced-choice. Instead of asking the man "Can you be depended upon—yes or no?", the test forces the man to choose between two alternatives: "I've got 'guts' "—"I can be depended upon." Items of this type were used in the Personal Inventory, long form, one of the most effective tests to date. Such a test is relatively free from the usual weakness that men will not tell the truth about themselves.

Personality tests, including forced-choice devices, are much improved if an empirical key is established. To develop such a key, the test is given to a large number of men, and ratings are later obtained to determine which men worked out well and which presented personality problems. The papers are analyzed to see what answers are given most often by the "problem" cases. Any new man who gives those answers is then identified as a potential problem and can be given special attention in processing. Jurgensen has applied the method in employee selection (**30**), using a technique quite applicable to military situations.

Projective tests are a third alternative requiring further investigation. Such tests present the man with an indefinite stimulus and require him to react to it. He may say what he sees in an inkblot, complete an unfinished sentence, make up a story about a dramatic picture, or the like. These tests have been very effective in clinical work; they evidently do tell a great deal about personality. No one has succeeded in applying them to military screening, but few reported studies have made a thorough attack upon the problem. This method, if it works, would be an ideal processing procedure, since the man reveals his personality through his actions rather than in a sometimes undependable self-description. Hunt and Stevenson suggest an empirical approach to this problem: for example, giving the inkblots to hundreds of men and noting what responses are given most often by good and by poor men (**29**). No one can say whether projective tests will work for screening neurotics, but one or two studies show sufficient promise to warrant a serious attempt to make them work (**2, 24, 37**).

The most relied-upon technique is the psychiatric interview, which attempts to elicit evidence of maladjustment in a short conversation (**49**). Although this method is known to work, there has been insufficient research on it. Studies should investigate how well different interviewers agree, how accurate their judgments are, what questions elicit the most useful information, and how psychiatric interviewers can best be trained. So important a problem is wide open for investigation. If 20 men were interviewed by different psychiatrists, would the psychiatrists agree about them? Are all psychiatric interviewers equally good judges of men? If not, by what tests can we select the best men for this responsibility? Recordings of interviews should show how effective interviewers proceed. If the process can be standardized, it would then be possible to train intelligent enlisted men to do at least preliminary screening. All such improvements would conserve the energies of the small number of qualified interviewers now available, and increase their effectiveness.

Assessment of the Total Personality

In contrast to the attempts to screen out misfits, recent personality studies have tried to come to an understanding of how personalities are organized. How a man exercises responsibility is no mere accident or outcome of training—it is the product of his childhood experiences, his affectional relations with his father, his feeling that other people approve of him, and other complex forces. Suitable techniques can determine such facts about the man, and with that knowledge it should be possible to assess his suitability for many types of duty. Will he vacillate in a crisis? Will he be lax in an attempt to be popular? Will he go stale on a job which does not bring excitement? Knowledge of this sort can be of great use.

The most impressive application of total assessment is found in the OSS (**52**). Here a three-day processing of men going into responsible assignments was conducted by a team of experts. By observing men in great variety of tests, social relations, and interviews, they were able to make useful estimates of their probable success in assignments. The neuropsychiatric breakdown rate for all the OSS was 0.26 percent; for men passed by the assessment team at Station S, only 0.04 percent (**52,** p. 433). The striking difference between this program and previous work on personality is that it attempts to determine the duties for which a man's strengths and weaknesses fit him, rather than merely to give him a pass or fail rating.

These investigators and others have found that one of the better indices of a man's probable adaptation in the field is the composite rating given him by his fellow assessees after several days of close association. This method of obtaining data should be given more widespread trial.

Assessment of the total personality is a new technique both in science and in military processing. It is difficult, but will be very valuable if it succeeds at all. It is most important for high level personnel, who can be given a variety of assignments, and whose success is important enough to warrant processing effort. Several studies are in progress which attempt to apply these techniques to business and professional people (for example, see **25**); parallel work on military men is essential.

Studies of Leadership

Another facet of the same problem is that no one now knows what sort of men make the best Naval officers. Moreover, no one has demonstrated that the man who is best as a junior officer makes the best commander or the best staff officer. Studies of the personality of effective and ineffective officers would be a major help in selecting new officers. Such a study should go much deeper than a few questionnaires. A thorough case study of several top-quality officers by a competent clinician would cast much light on the effective military personality. One approach to the problem is now in use by the American Institute of Research, studying Air Force officers and ONR scientists. This valuable technique tries to identify the *actions* which distinguish good and poor men. If this could be supplemented by studies of *personality structure* which leads to good or poor actions, one might be able to identify potential commanders even among seventeen-year-old recruits.

While the majority of the problems of leadership are discussed in another chapter, one problem is essentially a selection problem. That is the question as to whether a more efficient team can be assembled by matching the men to form a compatible group. It seems reasonable that every man would work better with some superiors and co-workers than others, but there are no studies on this point. There have been a few attempts to staff airplanes with groups of men who chose each other, and the same thing is done in an informal way in assembling submarine crews. Research on what sort of team brings out the best in each man would no doubt provide suggestions for placement of men and officers.

Criteria of Personality

A major problem in determining the extent to which personality problems arise in undersea warfare is the tradition against recording personality failures in service records. It is commendable for officers to avoid condemning a man for a crackup; a face-saving transfer or disqualification for medical reasons is a sound administrative practice. But some reliable method is needed to determine the incidence of psychiatric problems.

Since men passed by screening psychiatrists develop anxieties and tensions in the course of patrols, some method is needed to monitor personality in the field. Probably a procedure should be developed for a short psychiatric interview with every member of a submarine crew after every war patrol. It is unsound to assume that initial screening is a sufficient safeguard. Such further screening would identify problems about to come to the surface in erratic behavior and permit men to receive rest when they need it.

The Criterion

One of the greatest weaknesses of selection and training in the late war was the inadequacy of criteria. Without a suitable measure of how well men perform on the job, no one can say what tests are permitting good selection, which training methods are effective, and which men should be selected for added responsibility. Nearly every research report on a wartime selection program includes a comment that the research is inadequate because no fair criterion could be obtained.

A report from BuPers well demonstrates the seriousness of criterion errors. When poor tests were given to men graduating from Basic Engineering School, it seemed that prediction was poor, but that the Arithmetic test was the best test for prediction. When new and better achievement tests were designed and given to graduates, not only was validity raised, but it was found that a different set of tests (especially Mechanical Aptitude) selected the best men. The poor criterion, in other words, led to recommendation of the wrong selection battery (**44**, pp.

306–309). Poor criteria have an equally serious effect on training, since a poor method of evaluation will pass incompetent men to duty and perhaps fail good ones. Men may pass tests supposed to show proficiency, but if the tests themselves are invalid, months may pass before it is realized that the training program lacks effectiveness.

Any research that can be pursued now to develop good criteria in training schools or in the Fleet is desirable; even more desirable is research to develop good general methods for the design of criteria as new tasks become important.

Objective Criteria

One line of approach is to objectify criteria. A general opinion from a supervisor or observer is notoriously inadequate. Bias, "halo effect," differences between observers, and other errors, make such estimates crude at best. Methods of eliminating observer errors are many, but it is not certain which will be most practical.

One attack is to require superior officers to collect specific records of good and poor job performance, instead of merely rating from memory at the end of a six-month period. Such a procedure is now being developed for the Air Force by the American Institute for Research. The checkride in aviation is a technique based on specific observation rather than opinion; this method has been greatly improved by defining carefully the test maneuvers and the facts to be noted (**47**). Similar field tests of undersea warfare personnel can be used effectively, as demonstrated by the USS SYLPH, during the recent war. A few graduates from each class of the two fleet sound schools were sent to New London and tested at sea, and weaknesses which were detected led to revision of faulty programs in the schools. Perhaps all that can be said at present is that the weakness of subjective criteria has been overcome very little by research to date, but that studies of the type made during wartime should be pursued (**44**, pp. 357–379).

Combat Criteria

No research during peacetime can really evaluate a military selection or training program. The real test of a bombardier is whether his bombs hit the target even when he is under tension and threatened by flak. The test of a submarine officer's leadership is whether he can maintain the morale of a crew on a long patrol during which they are depth charged repeatedly. No method of making such tests is feasible in peacetime. And it is probably not true that the man who does best in simple peacetime demonstrations and dry runs will also have the reserve capacity which makes an outstanding combat performer. Some of the best men in wartime are those who were undistinguished in peacetime but were brought to a peak by a real challenge.

The problem, then, is to establish methods for evaluating how well men do in combat. These methods, if thought out in advance, should greatly increase the dependability of selection methods.

Multiple Criteria

Past workers have tried to predict a single criterion: "goodness" on the job. But performance is not so easily described; any job has many aspects, and most men are better at some aspects than others. Any research on prediction of complex duties should be so designed that a criterion will show how well the man does each aspect of his job. For example, the lookout might be characterized thus: "He is superior in locating small targets; he frequently misclassifies a cargo ship as a warship; he makes some mistakes in estimating heading, though he is generally accurate; he is one of the best men I've ever had for maintaining alertness on a long watch; and so on." Such a breakdown of the job should permit far more exact prediction of job performance than the usual single criterion. In aviation selection, personality tests predicted very poorly who would graduate from school (**18**); this is to be expected, because so many other factors condition success and failure in school that men with desirable personality makeup often fail, and men with poor personality makeup pull through. This is not to say that personality may be ignored. Some personality test might predict accurately a rating on "Annoys fellow crew members" or "Loses nerve in face of danger," and that prediction would be important.

Even if good multiple criteria were developed, no method is now available for handling them efficiently. Suppose there are 20 aspects to the job of operating a Loran transmitter; suppose there are tests which predict how well each man will do each aspect of the job. How can such data be used effectively in classification? It is important to develop the mathematics of prediction with multiple criteria. One suitable approach is factor analysis of the criteria.

This analysis may in turn have implications for billet analysis, if a job is found to break down into several psychologically-unrelated aspects (cf. **28,** pp. 20–24, 69–80). Speculating, it appears possible that we should find a widespread mathematical factor, running through such skills as determining a depth-charge setting, interpreting data from triangulation detectors, planning an interception course, etc. A man with superior mathematical ability might be better at all these functions than the person not so superior. Other shipboard functions might fall into other categories: listening jobs, visual jobs, jobs requiring mechanical insight. It might pay to consider redesigning billets so that the best mathematical minds were available for mathematical tasks, so the best ears were trained to do the auditory tasks from monitoring a torpedo to listening to the engine valves, and so on.

The most meaningful type of psychological measuring scale represents one thing and one thing only. A composite estimate which combines a man's knowledge of equipment, skill in detecting targets, speed in making repairs, alertness, and pleasantness, is a very poor basis for judging the man's fitness for a particular job. Training might be seriously deficient in one aspect, and yet all men might earn superior ratings at sea because of other assets. A man could be sent to sea lacking some one basic aptitude, yet this weakness might be ignored for years if his record reports only a general estimate of all-round quality. The only effective plan for assessing men is to measure separately each man's proficiency in each vital aspect of the job. The proper criterion for evaluating psychiatric screening is emotional stability on duty, not all-over proficiency. The criterion for evaluating target detection is the man's ability to pick up targets, not some other index such as the accuracy of bearings when he has a target. One of the few attempts to obtain multiple criteria for use in evaluating selection programs is to be credited to Navy aviation psychologists. They obtained nominations of good and poor fliers in carrier squadrons, and had the nominator tell why each man was good or poor. This led to a pool of men known to be very good in, say, emotional adequacy under pressure, and a second pool of men poor in this respect. A second pair of groups differing in teamwork or in flying skill was selected, and so on (**53**). The predictors could then be correlated with each aspect of the criterion. Final reports on this type of study are not available, but the approach is most promising and should be extended even though wartime criteria are lacking. Research can at least set up efficient plans for obtaining and using multiple criteria.

Qualitative Criteria

No measurement or rating gives as effective a picture of how a man works out on a job as does a description. This is especially true of officers, who can show a great variety of assets: daring, judgment, steadiness, resourcefulness, and so on. If we are ever to assess personal attributes, these assessments must be judged against a suitable description of performance on duty. One cannot predict a rating without knowing the man, the job to which he is assigned, and the man who will rate him. But perhaps emerging techniques can predict that he will (for example) "offer many original and useful suggestions to a friendly commanding officer, provided he does not have to take responsibility himself for deciding whether to adopt them." Qualitative criteria permit validating such predictions. Research on how to obtain and how to utilize qualitative criteria is required, since they fall outside the statistical tradition of previous research.

Training

Organization of Training

The general plan for training is frequently dictated by military convenience rather than by psychological factors. Even decisions to train or not to train men for a particular duty have been made without evidence. Wartime research often showed that needed training was not being provided because it was assumed that all men could do a job. For examples, see studies of night lookouts (**58,** pp. 63–78) and voice communication (**58,** pp. 98–105). Additional training often raises efficiency more cheaply than redesign of equipment. Research will show what skills can be improved by training.

Investigation of the overall organization of training is desirable, in order to consider such questions as the appropriate sequence of courses for particular duties. An example of such research led to a finding that men with navigator training learn in 35 hours radar bombardment techniques that other men require 85 hours to learn (**56**).

Another such question relates to the length of the training course. Whether five weeks or twelve are allowed to teach a specialty is largely a matter of chance, plus the ability of officers in different schools to present effective arguments. There is next to no evidence comparing the effectiveness of the best possible six-week course and the best possible eight-week course for gunners, and so on. One study of radar observers (bombardier) showed that 150 hours of additional careful training cut circular bombing error nearly in half (**58,** p. 18); but another study with Norden bombsights showed that, in a 25-mission training program, men gained nothing after the third mission (**16,** pp. 166–169). Such studies of various skills are indispensable to a rational plan which will provide each man with all the training he can profit from, without wasting his time in training exercises which produce no gain.

Refresher training has also rarely been organized systematically. It has been established repeatedly that men lose skills they do not practice. If a man operating sound gear has no opportunity to practice detection of mines, his skill may be below standard when it is needed. Refresher training is often given, but without any systematic basis. The man sent to school is the man whom the ship can spare, or who requests schooling; the man the ship relies on may not be sent. When an ASW vessel puts into port and operates for a few days with a "tame" submarine, the drills may be conned by a junior officer. When the ship puts to sea and makes a real contact, the commanding officer who skipped the refresher training often takes over the conn. The design of effective refresher training requires systematic study and trial. Where research would help is in development of evaluative procedures which could be applied to every man periodically to prove that his skills are intact. Such tests could be made at any shore installation, much as the CAA requires periodic tests of pilots. Ideally, a man would be permitted to perform a duty only if his most recent tests showed him to be fully qualified. The design of the best refresher training is itself a problem for research; the psychology of learning has developed much knowledge of how men learn new tasks, but very little of how they relearn what is partly forgotten.

A third program requiring extensive analysis is shipboard training. Partly this is a question of designing good devices (manuals, shipboard attack teachers, filmstrips, etc.) which can be used to duplicate school training. But there are special problems of shipboard training which have never been studied. Teaching methods which must function without a qualified instructor probably must be different from normal ones. How different, no one has determined.

Curricula and Methods

In designing any training, one must decide what to include. Some things can be learned at sea. Some theoretical topics can be included or omitted, at the option of the course planner. Some practical procedures, like making certain repairs, can be taught in school, or left in a manual to be looked up when they are needed. Whether these things should be placed in a curriculum is subject to research; in the past it has been left solely to judgment.

Use of Training Time

An effective training course is one in which every minute of the working day contributes to the man's later performance on duty. In war or peace, time cannot be wasted on ineffective lectures, on drills which actually confuse men at later stages of training, on films which entertain but do not change behavior, or on exercises which merely fill time when proper training facilities are not available for all men. Most training programs contain one or all of these dead spots. The straightforward research design would be to omit some of the topics or drill sessions, and to determine whether the omission made any noticeable difference in the man's performance, judged by any objective standard. Thus, a study of code learning showed that four hours drill per day taught as much as seven hours, saving three hours per man per day (**32**). An even more striking example of wasteful course planning is an Air Force study of gunners. When men who had taken the regular basic (ground) gunnery course were compared with men not so trained, there was no discernible difference in the groups after five air-training missions (**16**). In addition to studies of the training sessions themselves, studies should consider the effect of the total day's program during training. What is the maximum number of hours per day that can profitably be spent on balanced and varied training? What is the effect on course learning of effort put into calisthenics, policing barracks, etc.?

There is a shortcut, namely, the application of general principles of learning (**58,** pp. 124–126). Some of these are easily stated: Men forget a large amount of what they learn unless it is overlearned (practiced well beyond the point where they can just do the task) or unless they understand the reason why a certain performance is correct. Men who learn to perform an act under one set of conditions may not be able to perform it under other conditions. Some men find it harder to learn a skill than do others, and require more time. Spaced practice is better than massed practice for most jobs; six twenty-minute drills generally make a better program than two one-hour sessions. Knowledge that is tied together is effective knowledge; a course must be as unified as possible. Any training curriculum can be analyzed in terms of these and other known principles. Such analysis should lead to revision of present curricula, sharpening of theoretical problems requiring research, and development of general plans for future courses. Application of general principles must itself be checked by research; in code-voice training, how drills were distributed within a day was found to make no real difference despite theoretical expectations (**31**).

Instructional Methods

How a man should be taught can be answered partly from existing knowledge. Research on learning is extensive and definitely applies to naval training. It is known that men have trouble learning when they must learn many different words to express the same ideas, and this principle was applied to simplify the vocabulary of pilot training and submarine training (**1, 34**). Telling a man about something is much less effective than having him do it; this was applied repeatedly in the development of training drills. Any training drill must be realistic; it is usually worthless to teach a man a response if the gear he will have at sea calls for a different response, or gives him different signals. Knowledge of such principles as these is useful, but many training methods have failed to apply them adequately. Devices used during the war sometimes gave men just enough practice so that they felt confused, instead of continuing practice until they had grasped the right action. Some devices, on the contrary, wasted time in simple drills long after men understood the response required. A first step in evaluating the naval problem would be to review every specific training method to identify the principles violated.

But in many ways present principles are inadequate. We know a good deal about drill, but relatively little is known about teaching for *meaning*. How can we present a topic so that a man understands it? The psychology of explanation requires extensive study. The present principles—avoid complex terminology, use pictures for visual concepts, etc.—are but a limited guide to the man who must write a maintenance manual, a textbook, or a lecture.

Present principles tell what method of teaching is best on the average. Such studies, comparing the average learning of groups taught by different methods, are helpful, but they are inadequate. The method which is best on the average is not best for all men (**7, 58**). One study, for example, tested the valid rule that men should practice actively whatever they are supposed to learn. In learning the phonetic alphabet, active practice was found to be unimportant for men of high ability and men of high motivation, even though valuable for the average man. Probably if we checked any principle of learning, we would find that it applies to some men or to some types of learning more than others. If so, cost and time of training can be reduced by using the method appropriate for the men and topic to be taught. For any one purpose there is a best training method; but there is no one best teaching plan for all purposes and all men.

Attention should be called to the problem of the selection and training of instructors. This is a virtually unopened question. There is need for studies of instructor effectiveness, and of methods of recruiting and training instructors.

Training Aids and Devices

One of the most effective applications of research has been the development of synthetic trainers and visual aids. The evidence has been well-advertised that men learn better when they see films, or when they practice skills on trainers which duplicate tactical problems. These aids will be a permanent part of training. In the enthusiasm for aids, research has rarely been critical. Not every aid is magical; some of them are wasteful, and in some studies much less elaborate techniques have proved equally effective (**22, 35, 36**).

Because training devices are expensive and consume both production time and training time, they should never be employed except where they improve learning, and they should be so designed that every minute of time they require is effectively spent. A vast amount of research, well-financed, is required to establish the principles that make an effective training aid.

One way to proceed is by intensive analysis of present trainers. Studies of the echo-recognition and sound-recognition group trainers provide an example (**11, 13**). These devices were checked for conformity to psychological principles. Each single recorded drill selection was checked experimentally to determine if men succeeded in mastering it, if it was long enough to permit judgment and short enough to not waste time, if it was being presented the correct number of times for reinforcement, and so on. Tests were made for the trainers, so that weaknesses requiring review or revised training would be detected. Through such monitoring, the trainer was made more effective. Studies of a great variety of training devices should yield a body of experience as to sound design.

Wolfle has summarized the experience of those designing training aids in the last war (**58**, pp. 140–154). He demonstrates how application of principles of psychology improved these devices and stresses the importance of a research validation for every new trainer developed. Because training aids have their own technology, it is important for Wolfle's brief summary to be elaborated in a form available to all psychologists. Since synthetic trainers are little used except in the armed services, psychologists normally have no chance to develop know-how regarding their construction. They would be greatly aided by a thorough manual on the technology of trainers. Such a volume, prepared by engineers and psychologists, would draw on all available experience. It would not only list such general principles as "a scoring device is required," but would also discuss the types of scoring devices so far employed, their merits and limitations, whether it is desirable to have continuous scoring or periodic scoring, and other detailed questions. Such a thorough review of wartime experience would not only provide a guide to each new project, but would focus attention on questions which have not yet been settled experimentally. It would do much to eliminate the frequent experience of the last ten years, in which one group of scientists and engineers has laboriously and expensively reinvented a technique already in use somewhere else. Contracts should be let during peacetime to encourage groups of psychologists to work closely with designers and manufacturers in developing specific trainers.

In the design of training films, experimentation is required on many fundamental questions. Do questions inserted in a film improve its effectiveness? What types of pictures present ideas most clearly? What kind of learning is aided by pictures? Does it improve learning if an idea is woven into a dramatic story rather than presented straightforward? Is synchronized sound necessary for an effective film? What ideas can be presented as well by film-strip as by motion picture, or by silent picture as by sound picture? The answers to dozens of such questions, some of which are being attacked by the Instructional Film Research Project of Pennsylvania State College, will permit producers to make better films and will possibly lead to tremendous savings in production costs. Costs alone are not a major consideration, but reducing costs also increases the number of films that can be made. Since present films are made largely by traditional rules which may or may not be valid, large gains in learning impact are to be expected if current and future research can provide a scientific guide to making training films. In this connection, attention should be drawn to an excellent critical review of Navy films by Captain Exton, formerly with the Bureau of Naval Personnel (**15**, pp. 55–70). His dissatisfaction with the "Hollywood influence" is thoroughly justified, although the suggestions he makes for improving films have frequently not been established by research. Many effective studies may be made, testing one principle of film design at a time to determine how sound and important it is. Research should test present films in great detail, to learn what portions of them work and why. Other studies should make experimental films testing one question of film design at a time.

There are two competing points of view regarding training aids. One sees the aid as a master teacher, replacing the instructor. There are functions that self-sufficient training aids can perform, and research should seek to determine what the limits of such aids are. On the other hand, most aids are used as part of a curriculum which also includes lectures, drill sessions, motivating sessions, and other methods of instruction. Research on how to make a good film will not itself tell how to make a film which functions effectively in a total program. When considered as only one technique of teaching, one must ask, what can films (or other aids) do better than other instructional methods? Shall we use five minutes of film footage to motivate men to learn, or shall we leave that to the instructor and spend the costly footage on getting across complex ideas more thoroughly? How can films and trainers be most effectively used along with drills on actual equipment? Should all explanations and trainer drills precede drills at sea, or can training aids be used more effectively for review of difficult topics? These and other questions are open to research, but it is essential that studies go beyond demonstrating that training aids are good. They must, instead, find ways of making sure that the time and effort expended on training materials is fully productive. Too often, in the past, techniques have been adopted which did not produce enough gain to justify their cost in effort and money.

Research must also be directed to incidental aids. Do booklets, posters, and the like change men's behavior? Very little is known on this question, yet such aids occupied a great deal of manpower in wartime. Evaluation of incidental devices is not difficult. If it should be shown that posters and similar devices have any effect on behavior, research could then profitably turn to the question of how to design them for greatest potency.

Transfer of Training

Research on transfer has shown that what a man learns in one situation will transfer to a new situation if the two situations are similar, if the man understands a general rule applying to both situations, or if the original training employed a variety of stimuli. These principles were useful in designing military training, but they were effective in some problems and not in others. The difficulty is that present theory of transfer does not make clear just when situations are similar enough for transfer. So far, only experimental study of each new problem can determine for certain whether training will help. Under some circumstances, training may be positively harmful because the man is trained in one situation and fights in another. If a telephone talker learns to enunciate in a quiet room, he may be unintelligible when he relays messages on a noisy ship. Tracking exercises with a gun mounted on a classroom floor may lead to improvement; when the graduates try to track a target from a rolling deck, their training may prove helpful, unhelpful, or a source of confusion and interference. If the pips presented by a synthetic radar trainer do not exactly resemble signals from actual targets, men may learn to respond to false cues. Such problems can be reduced by training carefully designed so that the man responds to stimuli like those he will face on duty with responses like he will make on duty. When tests of the transfer value of the training are made, errors will be discovered and eliminated. But research which shows exactly how much similarity between training and duty is required, and just what aspects of similarity must be considered, would be especially helpful. Some studies have drawn attention to factors of this type (**26, 35, 36, 42, 48, 59, 61**).

If a training experience could transfer to a large number of situations, it would be most helpful. For example, it would be an advantage if training a man on one radar or one fighter plane would equip him to handle other models with only a brief check-out. This is in part a problem of equipment design. When the landing flap control is sometimes to the right of the pilot, sometimes to the left, serious accidents result because of negative transfer. All gear of a particular type should be as similar as possible: position and appearance of controls and indicators, speed of train and stiffness of controls, marking of reticles, and so on.

Error Analysis

For effective training it is necessary to determine the errors men make on the job and to remove them. Observing men as they perform their duties shows not only how competent they are, but reveals the specific errors they make. Similar studies of men at the end of training reveal their weaknesses. For example, it has been found that gunners tend to overestimate opening ranges (**27**); when such a systematic weakness is identified, specific training to correct the error can be introduced. Studies of operational difficulty (e.g., studies of relative difficulty of code characters (**41**)) show how much time should be spent on each response in the training course. Sometimes attention is drawn to aptitude factors to be considered in future selection. Occasionally, such studies lead to recommendations for modifying gear so that errors will be reduced.

Error analysis requires effective methods of evaluating performance. In all training schools, one of the major contributions of psychologists was their assistance in developing ways of measuring learning, by written tests, performance tests, and product ratings (**44**). In making these tests, they drew on every known principle of test design and developed a few new ones. The importance of these tests as a means of improving training cannot be overestimated. They are a sufficient reason for encouraging research on test design.

It is difficult to point out types of research needed in this area, since testing methods are somewhat specific to particular situations. Performance tests are especially important, however, since they have received less attention than their validity warrants. How a man actually does a job can be evaluated by an observer, by automatic scoring devices such as a phototube indicator of time on target, by motion pictures of his performance, and by other methods. It would be helpful to have a comparison of methods, such as the study of checksight scoring methods (**6**), as a guide to future performance measures.

SUMMARY

Selection research has increased the effectiveness of military forces by placing superior men in responsible posts. Selection and placement of men can be improved by mathematical studies of mass processing, of the multiple cutoff method, and of factor analysis. Billets should be studied in terms of their human requirements, and the significant human aptitudes should be classified. Test design should be deliberate, based on principles of measurement, rather than a trial-and-error process. Basic studies of test design are much needed. Improved technical methods of classifying men must be supplemented by careful use of the methods.

A frontier barely explored is personality assessment. Tests of personality can be improved. Studies should consider how normal men function, rather than identifying potential breakdown cases only.

Research has been handicapped by inadequate criteria of proficiency. Procedures for judging the effectiveness of men on the job can be developed.

Training has frequently been bettered by criticism in the light of psychological principles. Specific studies of local training programs have led to alterations of course content, length, or methods. Further studies of specific courses are advisable, but research setting up broad principles is essential. Time allotments, explanation procedures, practice conditions, and the design of training aids are promising subjects for research.

References

1. Abrams, M. H., Neff, W. D., et al. *A project for standardizing submarine phraseology and developing a training program in submarine voice communication* OSRD 4795, OEMsr-658, OEMsr-830, and OEMsr-1128, Service Projects N-118 and N-109, Div. 6, Report P 57/1421, Harvard PAL, CUDWR, and The Psychological Corp. New London, Feb. 28, 1945.
2. Abt, L. E. The efficiency of the group Rorschach test in the psychiatric screening of Marine Corps recruits. *J. Psychol.*, 1947, **23**, 205–218.
3. Bartlett, N. R. *An investigation of the Billet Qualifications Blank (Navpers16418) for its suitability with combat-experienced candidates for submarine training.* Bur. Med. Surg., Research Project No. X-598 (Sub. No. 129). New London, Conn.: Med. Resch. Dept., U. S. Submarine Base, January 14, 1946.
4. Bennett, G. K., & Doppelt, J. E. The evaluation of pairs of tests for guidance use. *Educ. psychol. Measmt.*, 1948, **8**, 319–326.
5. Biel, W. C., & Searle, L. V. *A study of the Tufts Tracking Trainer as a selection and training device for trackers on the director, M-7.* OSRD 3606, OEMsr-581, Service Project SOS-6, Report 12, Tufts College. Medford, Mass., April 24, 1944.
6. Birmingham, H. P., Householder, A. S., Kappauf, W. E., & Taylor, F. V. *Check sight scoring methods.* OSRD 4525, OEMsr-815, Service Project N-111, Report 1, Brown Univ. Providence, R.I., Jan. 3, 1945.
7. Bond, A. D. *An experiment in the teaching of genetics.* Teachers College, Columbia Univ., 1940.
8. Brogden, H. E. An approach to the problem of differential prediction. *Psychometrika*, 1946, **11**, 139–154.
9. Conrad, H. S. *Characteristics and uses of item-analysis data.* OSRD 4034, Project N-11, Report 11, CEEB. Princeton, N. J., Aug. 19, 1944.
10. Conrad, H. S., & Satter, G. A. *The use of test scores and quality-classification ratings in predicting success in Electrician's Mates school.* OSRD 5667, Project N-106, Report 23, CEEB. Princeton, N. J., Sept. 13, 1945.
11. Cronbach, L. J. *Analysis of echo-recognition training in the light of certain psychological principles.* Univ. of Calif. Div. of War Research. San Diego, July 15, 1944.
12. Cronbach, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, **6**, 475–494.
13. Cronbach, L. J., & Lovell, D. F. *Learning studies on the Sound Recognition Group Trainer.* Reports M 319, 338, 352, 351, 359, 371. Univ. of Calif. Div. of War Research. San Diego, 1945.
14. Duff, Ivan F. *Medical study of the experiences of submariners as recorded in 1471 submarine patrol reports in World War II.* Research Division, Bur. Med. Surg., Washington, D. C., (undated), 269–279.
15. Exton, W., Jr. *Audiovisual aids to instruction.* New York: McGraw-Hill, 1947.
16. Flanagan, J. C. (Ed.) *The aviation psychology program in the army air forces.* AAF Aviation Psychol. Prog. Res. Rep. No. 1. Washington, D. C.: Govt. Printing Off., 1948.
17. Gibson, J. J. (Ed.) *Motion picture testing and research.* AAF Aviation Psychol. Prog. Res. Rep. No. 7. Washington, D. C.: Govt. Printing Off., 1947.
18. Guilford, J. P. (Ed.) *Printed classification tests.* AAF Aviation Psychol. Prog. Res. Rep. No. 5. Washington, D. C.: Govt. Printing Off., 1947.
19. Guilford, J. P. Factor analysis in a test-development program. *Psychol. Rev.*, 1948, **55**, 79–94.
20. Gulliksen, H. *Selection of test items by correlation with an external criterion, as applied to the Mechanical Comprehension Test—OQT-0-2.* OSRD 3187, Project N-106, Memo. 12, CEEB. Princeton, N. J., Jan. 8, 1944.
21. Gulliksen, H. The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 1945, **10**, 79–91.
22. Hall, W. E., & Cushing, J. The relative value of three methods of presenting learning materials. *J. Psychol.*, 1947, **24**, 57–62.
23. Harris, J. D. *Report on studies in group audiometry on submarine personnel.* Bur. Med. Surg., Research Project X-328 (Sub. No. 57). New London, Conn.: Med. Resch. Dept., U. S. Submarine Base, Nov. 15, 1944.
24. Harrower-Erickson, M. R. The value and limitations of the so-called "neurotic signs." *Rorschach Res. Exch.*, 1942, **6**, 109–114.
25. Henry, W. E. The business executive: a study in the psycho-dynamics of a social role *Amer. J. Sociol.*, 1949, **54**, 286–291.
26. Herbert, M. J., & Keller, F. S. *The use of hand-sent clear text as practice material in learning to receive Morse code.* OSRD 5168, OEMsr-830, Service Project SC-88, Report 4, The Psychological Corp. New York, June 4, 1945.

27. Horowitz, M. W., & Kappauf, W. E. *Aerial target range estimation.* OSRD 5301, OEMsr-700, Service Project N-105, Report 4, Brown Univ. Providence, R. I., July 3, 1945.
28. Horst, P. *Prediction of personal adjustment.* Bulletin No. 48, 1941. New York: Social Science Research Council.
29. Hunt, W. A., & Stevenson, I. Psychological testing in military clinical psychology: II. Personality testing. *Psychol. Rev.*, 1946, **53,** 107–115.
30. Jurgensen, C. E. Report on the "Classification Inventory," a personality test for industrial use. *J. appl. Psychol.*, 1944, **28,** 445–460.
31. Keller, F. S., & Estes, K. W. *Distribution of practice in code learning.* OSRD 4330, OEMsr-830, Service Project SC-88, Report 2, The Psychological Corp. New York, Nov. 10, 1944.
32. Keller, F. S., & Estes, K. W. *The relative effectiveness of four and seven hours of daily code practice.* OSRD 4750, OEMsr-830, Service Project SC-88, Report 3, The Psychological Corp. New York, Feb. 26, 1945.
33. Keller, M. F., & Jerome, E. A. *Progress in receving International Morse code.* OSRD 5366, OEMsr-830, Service Project SC-88, Report 5, The Psychological Corp. New York, July 23, 1945.
34. Kelly, E. L. *The flight instructor's vocabulary.* CAA, Division of Research, Report No. 22. Washington, D. C., Oct. 1943.
35. Knauft, E. B., Spragg, S. D. S., & Taylor, H. A. *The design, construction, and validation of a mechanical radar pip-matching trainer.* OSRD 4636A, Service Project N-114, Report 16, Univ. of Wisconsin. Madison, Feb. 8, 1945.
36. Knauft, E. B., Spragg, S. D. S., & Taylor, H. A. *Validation of the Modified Army BC-968-A Radar Trainer.* OSRD 4476, OEMsr-1171, Service Project N-114, Report 15, Univ. of Wisconsin. Madison, Feb. 8, 1945.
37. Munroe, R. L. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. *Appl. Psychol. Monogr.*, 1945, No. 7.
38. Peterson, D. A. *Classification data available to ships' officers.* OSRD 5145, Project N-106, Memo. No. 14, CEEB. Princeton, N. J., May 30, 1945.
39. Prentice, W. C. H. *A study of the performance of night lookouts aboard ship.* OSRD 4087, OEMsr-1136, Service Project N-115, Report 2, Princeton Univ. Princeton, N. J. Oct. 15, 1944.
40. Ruch, F. L. A comparative study of the predictive efficiency of batteries of tests selected by the Wherry-Doolittle and a multiple-cutting score method. *Amer. Psychologist*, 1948, **3,** 291.
41. Seashore, H. G., & Kurtz, A. K. *Analysis of errors in copying code.* OSRD 4010, OEMsr-830, Service Project N-107, Report 5, The Psychological Corp. New York, Aug. 12, 1944.
42. Seashore, H. G., & Stuntz, S. *Training radio operators to copy code through interference.* OSRD 3747, OEMsr-830, Service Project N-107, Report 4, The Psychological Corp. New York, June 7, 1944.
43. Springbett, B. M. A method of obtaining critical scores from successive residues. *Bull. Canad. psychol. Assoc.*, 1944, **4,** 28–30.
44. Stuit, D. B. (Ed.) *Personnel research and test development in the bureau of naval personnel.* Princeton, N. J.: Princeton Univ. Press, 1947.
45. Taylor, H. C., & Russell, J. T. The relationship of validity coefficients to the practical effectiveness of tests in selection; discussion and tables. *J. appl. Psychol.*, 1939, **23,** 565–578.
46. Thorndike, R. L. (Ed.) *Research problems and techniques.* AAF Aviation Psychol. Res. Rep., No. 3. Washington, D. C.: Govt. Printing Off., 1947.
47. Viteles, M. S. The aircraft pilot: five years of research; a summary of outcomes. *Psychol. Bull.*, 1945, **42,** 489–526.
48. Viteles, M. S., Gorsuch, J. H., et al. *An investigation of the Range Estimation Trainer, Device 5C-4, as a method of teaching range estimation.* OSRD 4263, OEMsr-700, Service Project N-105, Memo. 11, Univ. of Pennsylvania. Philadelphia, Oct. 18, 1944.
49. Willmon, T. L. *Consideration of submarine medical problems and current status of investigations at U. S. Naval Medical Research. Laboratory.* New London, Conn.: Med. Resch. Dept., U. S. Submarine Base, Feb. 21 1948.
50. Wulfeck, W. H. Learning the two-hand coordination test. *J. appl. Psychol.*, 1942, **26** 41–49.
51. Wyatt, R. F. Improvability of pitch discrimination. *Psychol. Monogr.*, 1945, **58,** No. 2.
52. *Assessment of men.* OSS Assessment Staff. New York: Rinehart, 1948.

53. *The combat criterion program. Aviation Psychol. Tech. Memo.*, No. 4, Dec. 1, 1944. Washington, D. C.: Div. Aviat. Med., Bur. Med. Surg., Navy Dept.

54. *A comparative study of verbalized and projected pictorial tests in gunnery.* Unpublished. Training Aids Section, 9th Naval District Headquarters, Great Lakes, Ill., 1945.

55. *Construction of keys X, Y, and Z for the Biographical Inventory. Aviation Psychol. Tech. Memo.*, No. 3, Nov. 1, 1944. Washington, D. C.: Div. Aviat. Med., Bur. Med. Surg., Navy Dept.

56. *Final report on extended training experiment.* Headquarters AAF, Victorville, Calif., July 11, 1945, prepared by Service Projects SC-70 and NS-146 in collaboration with personnel of AAF, Victorville, Calif.

57. *Human factors in military efficiency. Vol. I. Aptitude and classification.* Summary Technical Report of the Applied Psychology Panel, NDRC. Washington, D. C.: OSRD, 1946.

58. *Human factors in military efficiency. Vol. II. Training and equipment.* Summary Technical Report of the Applied Psychology Panel, NDRC. Washington, D. C.: OSRD, 1946.

59. *Studies in aided tracking.* Progress Report to the Services 102, Div. 7, Fire Control, October, 1945.

60. *Submarines—Enlisted billet descriptions (Navpers 15646).* Bur. Nav. Pers., Washington, 1946.

61. *A study of factors determining accuracy of tracking by means of handwheel control.* OSRD 3451, OEMsr-453, Div. 7, NDRC Section D-2, Fire Control, Progress Report to the Services 71, Foxboro Company, Sept. 30, 1942.

62. *Value of an audience-participation technique in film-strip presentation.* Inform. and Educ. Div. Rep. No. B-154, HQ, Army Service Forces, Washington, D. C., June 30, 1945.

PART IX.

PERSONNEL RESOURCES

Chapter 24

PERSONNEL RESOURCES FOR RESEARCH IN APPLIED EXPERIMENTAL PSYCHOLOGY

JOHN L. KENNEDY

Tufts College

AND

RUTH CRUIKSHANK BUSSEY

Washington, D. C.

I. Introduction

At the present time, any large-scale program for work in experimental psychology applied to the field of engineering design of equipment for most efficient human use will be severely handicapped by lack of trained personnel at all levels. Those who have been associated with the staffing of laboratories and field stations for this work will testify that the important considerations are not money, time, or facilities, but who can be found to do the experimental work. It will be the purpose of this paper to state some facts concerning the current manpower shortage in this specialty and to recommend to those who are requesting further research work in applied experimental psychology a feasible program for insuring a steady supply of trained personnel.

II. Manpower estimate

With the help of the records of the American Psychological Association Office in Washington, D. C. and Drs. Dael Wolfle, Executive Secretary of the APA, and Helen Wolfle, Dr. Bussey has prepared Tables I and II, showing the best estimates available of the number of members and associates of the APA who might have the training and interest to work effectively. Since this survey is based upon the individual's own statements as to his research and teaching interests rather than on some more objective criterion, it is probably true that the frequencies are actually too high.

The figures of Table I are based upon information collected in 1945 and 1946. Table I shows that a total of 1084 members and associate members of the APA have expressed interests in teaching or research which might fit them for work in some phase of applied experimental psychology.

In Table II it is seen that, in 1945, approximately 26% of the membership of the Association would possibly have been available, by training and interest, for military applied experimental psychology. An estimate based upon a proportionate increase in number of such individuals to the growth of the Association (4183 in 1945 to 5009 in 1947) yields the number 1245, of possible personnel to staff military projects. The increase from 1084 to 1245 is probably spurious, since the emphasis in recent years has been on clinical rather than experimental training and the new members of the Association have been drawn largely from the clinical group.

But, assuming that 1084 of the membership of the Association represents a reasonable figure for the number of experimentally-trained psychologists, it is obvious that this figure should be reduced by an unknown number of persons who would be uninterested in applied experimental psychology. The full-time teaching staffs of departments of psychology in universities and colleges might eliminate as high as 90% of the 1084. Other considerations such as age and sex would also operate to reduce this number still further. Although the available data are scanty, the conclusion seems justified that not more than 100 fully qualified psychologists are now available in this country to carry on the kind of research and development on military problems which proved so profitable to the services during World War II.

TABLE I

NUMBER OF MEMBERS AND ASSOCIATE MEMBERS OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION, AS OF 1945, EXPRESSING RESEARCH OR TEACHING INTERESTS IN THE FIELD OF EXPERIMENTAL PSYCHOLOGY OR ALLIED AREAS

Expressed Interest*	Number		
	Men	Women	Total
Experimental	572	114	686
Perception (vision, etc.)	33	5	38
Physiological	81	20	101
Comparative or animal	42	16	58
Aviation psychology	31	1	32
Psychomotor activity	10	0	10
Psychophysics	2	0	2
Fellow or associate in Theoretical-Experimental section†	109	24	133
Estimated number from those in war work‡	23	1	24
Total	903	181	1084

* As indicated in 1945 yearbook. Each member is listed only once. Listing is checked in order of appearance in table. Thus, if interest is both experimental and physiological, the member is counted under experimental. If comparative is listed and the member is also a member of the Theoretical-Experimental Division of the APA (1946) roster, the member is counted under comparative.

† The 1946 roster of fellows or associates who belong to the Division of Theoretical-Experimental Psychology included 452 people, of whom 319 expressed interest in at least one of the experimental psychology categories. However, 133 members of the division had not otherwise expressed their interest in experimental psychology.

‡ A total of 126 men and 6 women who indicated military or civilian war work did not indicate either research or teaching interest. A check on a limited sample of these persons (Part I, questionnaire for 1948 yearbook) indicated that about 2 out of 7 of these are interested in experimental psychology.

TABLE II

PERCENTAGE OF TOTAL APA MEMBERSHIP OF 1945 INTERESTED IN EXPERIMENTAL PSYCHOLOGY AND ESTIMATED NUMBER OF APA MEMBERS OF 1947 INTERESTED IN EXPERIMENTAL PSYCHOLOGY

Group	Year	
	1945	1947
Total membership	4183	5009*
Foreign affiliates	54	65
Interest in experimental psychology†	1084	1245‡
Percentage of membership interested in experimental psychology	26.2%	26.2%

* Figure from APA office, August 1947.

† Not including foreign affiliates.

‡ Estimated on the basis of 1945 count. This is, however, probably larger than would be obtained from a count similar to that of 1945 if such data were available. The increase in membership is among the more recently trained personnel whose practical interests are in line with the needs of the present time and is accordingly undoubtedly higher in the fields of applied and clinical psychology.

III. RECOMMENDATION

It appears to the present writer that the military applications of experimental psychology will be slow in development, if a special training program is not planned and executed. The training of an experimental psychologist is expensive and time consuming. The universities probably cannot afford to initiate such training without substantial subsidy from industry and government.

As a stop-gap measure, it is recommended that part of the funds of the Committee on Undersea Warfare might be earmarked for use in supporting training programs in applied experimental psychology at a few selected institutions whose present facilities might be expanded to provide this kind of training with the minimum of expense.

NAME INDEX

Page numbers are given both for mentions of authors' names and for numerical references to their publications that are listed at the ends of chapters.

SUBJECT INDEX

Consecutive passages extending over more than one page are indexed by the first page only.