Proceedings: International Symposium on Military Psychology (1st : 1957 : Brussels) (1961)

Pages
243

Size
7 x 10

ISBN
0309363446

Frank A. Geldard and Marilyn C. Lee, Editors; Division of Anthropology and Psychology; National Research Council

PROCEEDINGS

First International Symposium on Military Psychology

Edited by Frank A. Geldard and Marilyn C. Lee

Organized by the Division of Anthropology and Psychology
National Academy of Sciences-National Research Council
under the auspices of Headquarters, Air Research and
Development Command, United States Air Force

26–27 JULY 1957

BRUSSELS, BELGIUM

Publication 894

National Academy of Sciences-National Research Council, Washington, D.C.

1961

Library of Congress Catalog Card Number 60–60082

TO

COLONEL PHILIP H. MITCHELL, USAF

whose guiding hand was everywhere evident in that which follows

Preface

This volume presents the Proceedings of the First International Symposium on Military Psychology, held at the Palais des Académies, Brussels, Belgium, on July 26-27, 1957. The Symposium was arranged to take place immediately prior to the 15th International Congress of Psychology, which was convened at the University of Brussels July 28 through August 3 of the same year. Sponsored by the U. S. Air Force Air Research and Development Command in response to many requests for such a meeting, the conference was organized by the Committee on International Relations in Psychology of the Division of Anthropology and Psychology, of the U. S. National Academy of Sciences—National Research Council. General arrangements were under the direction of Dr. Glen Finch, Executive Secretary of the Division.

The Symposium represents a major contribution to international scientific understanding and cooperation. It was attended by about two hundred psychologists from 15 European countries, the Near East, Canada, and the United States, who heard and participated in the discussion of sixteen invited papers covering nearly all major areas of military psychology. General topics explored included: manpower analysis, selection and classification of personnel, military training and education, psychophysiology, human engineering, proficiency measurement, and military management and morale. An evening address on military psychology in the United States by Dr. Arthur W. Melton, then Director of the U. S. Air Force Personnel and Training Research Center, rounded out the extensive program presented to the participants.

Planning for the Symposium was begun in February, 1956, when the NAS-NRC Committed on International Relations in Psychology met under the chairmanship of Dr. Herbert S. Langfeld to discuss with invited Service representatives the possibility of bringing together military psychologists from various countries to discuss problems of mutual interest. At this meeting agreement as to the value and feasibility of the plan was reached, and arrangements were made for its joint implementation by the Academy-Research Council and the Air Force.

It was also decided that the cooperation of an advisory committee composed of European military psychologists should be sought in connection with the specific planning of the Symposium. Since several members of the proposed committee were due to gather in October at Strasbourg, France, for the initial meeting of the then recently created Section of Experimental Psychology and Animal Behavior of the International Union of Biological Sciences, advantage was taken of this meeting to hold an international planning session for the Symposium. This took place at the Maison Rouge in Strasbourg on October 3, 1956, with fourteen military psychologists attending from Belgium, Denmark, France, Western Germany, Great Britain, Italy, the Netherlands, Norway, Sweden, Switzerland, and the United States. Representatives from Finland and Spain were added later to the group.

The Symposium was thus brought into potential being, its subsequent realization being achieved after several intervening months of cooperative activity.

Directing the arrangements in Europe was the Symposium's general chairman, Dr. Frank A. Geldard of the University of Virginia, who was on leave from his academic duties during 1956-57 to serve as Scientific Liaison Officer with the London Branch of the U.S. Office of Naval Research, and was thus in a strategic position to coordinate activities. Cooperating in the preparations were: Dr. Louis Delys, Secretary-General of the 15th International Congress of Psychology, who helped to integrate the planning of the Symposium with that of the Congress; Dr. Marilyn C. Lee of the USAF Office of Scientific Research, who handled many of the arrangements in Brussels; Dr. Clifford Frisby, Director of the National Institute of Industrial Psychology of Great Britain, who coordinated the contributions from the United Kingdom and gave freely of his time and energies in other ways; and Lt. Col. Charles Chandessais, Director of Human Sciences (Commission des Sciences de l'Homme) of the Permanent Secretariat of the Scientific Action Committee of National Defense (Secrétariat Permanent du Comité d'Action Scientifique de Défense Nationale), who performed similar services in connection with French participation.

The planning of the program was guided by the principle that geographic representation should be as inclusive as possible, and that an attempt should be made to insure participation by all countries of Western Europe having researchers active in the field of military psychology. The final program embodied this aim: of the sixteen papers presented, there were four from the United Kingdom, two from France, and one each from Belgium, Finland, Italy, the Netherlands, Norway, Spain, and Sweden. The official discussants of the papers represented military psychology in Austria, Denmark, Finland, France, Western Germany, Italy, the Netherlands, Spain, Switzerland, and the United Kingdom. The United States contributed three papers and four discussants. The chairmen of the five Symposium sessions were from Canada, France, United States, the Netherlands, and the United Kingdom, and the attending delegates represented all the above mentioned countries plus the Union of South Africa, Egypt, Greece, Hungary, Israel, Jugoslavia, and Turkey. There can be little doubt that the Symposium was truly international in character.

The conduct of the meetings was as follows. All papers were pre-published in English and French and distributed to the participants for study prior to the conference. At the Symposium itself the authors amplified or illustrated their written contributions. Discussion of each paper was begun with a commentary by a previously designated discussant, and then turned over to the floor. The meetings were conducted in French and English, Simultaneous translation facilities being provided. All the proceedings of the Symposium were recorded on tape and subsequently transcribed.

The present volume presents all seventeen Symposium communications, together with the commentaries of the discussants and digests of the subsequent discussions. Any new material presented in the form of introductory remarks by the authors has been integrated with the original papers. All the papers are given in English and summarized in both English and French; remarks offered during the discussions by French-speaking discussants and participants are presented in English translation. The organization is chronological, and strives to represent the proceedings of the Symposium as nearly as possible as they actually occurred.

The steps prerequisite to the accomplishment of the aims set forth in the foregoing paragraph are very considerable ones, demanding much in energy and skill, alike. It is a pleasure for the senior editor to acknowledge the prodigious amount of work invested and rare judgment displayed by the junior editor, Dr. Marilyn C. Lee, in completing the ground work necessary to the readying of the manuscript of

these Proceedings for publication. Thanks to her great linguistic skill, her facility in English and French (as well as American!), she was able to blend discussion, introductory statements, and pre-published papers into a meaningful whole. Successful syntheses of this kind come only out of wisdom and judicious care.

It is a genuine pleasure to acknowledge the gracious and tactful supervision given by the chairmen of the four scientific sessions: Doctors Myers, van der Giessen, and Wilson and Lt. Col. Chandessais. All unfailingly kept the discussions alert and "on the track"; all provided just the right transitions when they were needed.

It was a matter of great regret that Professor Herbert S. Langfeld, who had been so influential in the development of plans for the Symposium, was himself prevented, by doctor's orders, from making the ocean crossing to be present at the Brussels meeting. In the course of the sessions, the General Chairman appointed a committee consisting of Dr. Frisby, Col. Chandessais, and Dr. Miles as chairman, to present a suitable resolution expressing the appreciation of the Symposium members to Professor Langfeld and their regret at his inability to participate. The committee cabled Dr. Langfeld, as his Bar Harbor summer home, as follows: "The realized success of this International Symposium on Military Psychology owes much to your background preparations and foresight. The Symposium regrets your absence and extends its best wishes." It is known that this message was a source of real satisfaction to Professor Langfeld; the Symposium represented the last of a long line of persistent and unselfish efforts on his part to improve international relations in psychology.

Frank A. Geldard
General Chairman
First International Symposium on
Military Psychology

Charlottesville, Virginia

CONTENTS

SESSION I

Session Chairman: Dr. C. Roger Myers

Department of Psychology, University of Toronto

Toronto, Canada

Remarks of Welcome

Dr. Frank A. Geldard, General Chairman

Scientific Liaison Officer (Psychology), U. S. Office of Naval Research

London, England

Baron P. Jaques de Dixmude

Committee Chairman, Belgian Armed Forces Chief of Staff

Col. Philip H. Mitchell

Director, Human Factors Division

U. S. Air Force Air Research and Development Command

Dr. Frank A. Geldard:

Ladies and gentlemen, no general chairman of an international conference ever faced a more friendly audience. To my old friends from America I express my greetings and the hope that your long journey will be in every way rewarding. To my new friends from Europe I wish both to extend greetings and to express my appreciation for the heart-warming hospitality experienced during the travels of the past year. Everywhere my family and I have been—from Scotland to Italy, from Spain to Finland—we have been received with the greatest cordiality. Probably there has never been anything military which has proved to be more agreeable or congenial than the year spent with the London Branch of the Office of Naval Research.

Having lived with the International Symposium on Military Psychology since its conception nine months ago at Strasbourg, and having watched it develop through the fetal and embryonic stages, there now remains very little for me to do beyond calling the attending physicians and pacing the corridors during birth. This I do by first asking Baron Jacques de Dixmude of the Belgian Armed Forces to extend a word of greeting on behalf of the host country.

Le baron P. Jacques de Dixmude:

Mesdames, Messieurs,

Je suis particulièrement heureux de vous souhaiter la bien-venue en Belgique à ce Symposium sur la Psychologie Militaire. Comme vous êtes tous des psychologues avertis, vous aurez déjà deviné les raisons pour lesquelles ma satisfaction est si grande. Il y a d'abord le fait que c'est la première fois, à ma connaissance, que se trouve en Europe une réunion internationale des personnalités spécialisées en psychologie militaire.

Permettez-moi de souligner immediatement que nous devons cette initiative à nos amis Américains. Aussi, le remerciement s'adresse au Dr. Glen Finch, Secretaire Exécutif de la Division d'Anthropologie et Psychologie de l'Academie Nationale des Sciences—le Conseil National de Recherche des États-Unis (Executive Secretary of the Division of Anthropology and Psychology of the U. S. National Academy of Sciences—National Research Council), à M. le professeur Frank Geldard, Agent Scientifique de Liaison (Psychologie) du Bureau de Recherche de la Marine des États-Unis (Scientific Liaison Officer [Psychology] of the U. S. Office of Naval Research), et à tous ceux qui ont contribué au succès de cette réunion.

Il est un fait que nous avons le droit de constater, que si la recherche scientifique est permanente et continue, elle progresse par bonds particulièrement remarquables chaque fois que la guerre met les hommes aux prises les uns avec les autres, exacerbant toutes leurs facultés pour la gagner. Mais entre les guerres, les recherches sur les aptitudes et les réactions de l'homme dans un éventuel conflit futur ne sont plus conduites que par quelques psychologues spécialisés, cloisonnés dans les frontières de leur pays. C'est pourquoi je suis si heureux de vous voir

aujourd'hui réunis pour exposer le fruit de vos travaux, confronter vos vues, et surtout pour faire bénéficier les autres de votre science et de votre expérience. Permettez-moi d'espérer que vous ne vous séparerez pas sans avoir organisé les bases d'une recherche coordonnée et d'une collaboration toujours plus efficace.

En cette époque passionnante—où la science met à la disposition des militaires des armes d'une puissance destructive sans précédent, des appareillages d'une complexité toujours plus étonnante, des avions dont les performances n'ont pour limite que les réactions des pilotes—il n'est pas douteux que la connaissance de l'homme soit plus que jamais nécessaire, et qu'il faut qu'il soit étudié isolément et en groupe. De plus, au moment où la plupart des pays donnent des signes de fatigue dans l'effort qu'ils s'imposent pour assurer une défense commune, il est plus que jamais nécessaire d'utiliser les hommes au maximum de leurs possibilités, de les instruire toujours mieux et plus vite, et surtout de les armer de ces forces morales qui, en définitive, garantissent la victoire finale.

Enfin, Messieurs, vous comprendrez ma joie et ma fierté que ce Symposium ait lieu à Bruxelles. La Belgique est un petit pays, mais elle est au carrefour des civilisations que nous avons à défendre.

Je vous souhaite, Mesdames et Messieurs, de retirer le maximum de satisfaction de votre séjour dans notre capitale.

Summary

The Symposium representatives are greeted and thanks are expressed to the American initiators of the conference. The importance of sharing the results and experience of military psychologists in different countries is emphasized, and hope is expressed that the Symposium will help to lay the foundations for coordinated research. The speaker notes the increasing necessity, in an era of advancing technology and international tension, of psychological research aimed at maximizing the utilization of human capacities.

Col. Philip H. Mitchell:

Thank you, Baron Dixmude. I am sure I express the feelings of all the Symposium participants when I say that we are indeed grateful for the opportunity of meeting here in your charming and beautiful city.

I should like at this time to enumerate the goals which motivated the U. S. Air Force in its sponsorship of this Symposium.

First, it was hoped through the conference to facilitate the exchange of scientific information among our respective countries. It is believed that such an exchange of knowledge and ideas will advance technological progress and prove of mutual benefit to all Symposium participants.

Second, the Air Force as a social institution has a major responsibility for furthering the growth and expansion of military psychology in particular, and of related social and biological sciences in general. Sponsorship of the present Symposium represents one way in which we are attempting to fulfill this obligation.

Third, we wished to provide American behavioral scientists with an opportunity to learn more about European psychology. Previous visits by individual American scientists to these shores have proved so beneficial that Air Force psychologists have urged for some time that a more systematic arrangement be effected for meeting with their European colleagues. This Symposium, organized by the U. S. National Academy of Sciences—National Research Council, and subsidized by the USAF Air Research and Development Command, is the result of those requests. The U. S. Air Force is proud to have had a part in making possible so distinguished an assembly.

The great scientific traditions of the European continent provide a most appropriate background for the discussions to follow. It may be recalled that it was out of this culture that there arose one of the fundamental tenets of Western thought—belief in the value of the individual and his right to develop to his full potential. Much of the military research on problems of personnel classification, safety and survival, training methods, and human engineering can be said to be guided by this belief in the importance of the individual. It is to be hoped that this credo will be furthered by the sharing of ideas and improvement of understanding to which we shall be dedicating our efforts at this Symposium.

594869 O - 61 - 2

APPROACHES TO THE PROBLEM OF FLYING SAFETY

Vidkunn Coucheron-Jarl, Rolf Gerhardt, and Erik Riis
Psychological Division, Norwegian Armed Forces
Oslo, Norway

With present methods there seems to be little possibility of improving flying safety through increasing the efficiency of pre-selection of pilot candidates. It is apparent that multiple correlations of psychological selection batteries with training criteria have reached a barrier at about .60. Thus, new leads, both in selection procedures and in general preventive measures, must be sought through detailed and systematic study of the performance of pilots in service.

The traditional administrative approach to the study of flying safety has been the analysis of aircraft accidents.

An accident is an unforseen occurrence, the term being used especially to refer to an unexpected event of an undesirable kind. In the case of flying accidents, the criterion has been the degree of damage to the aircraft and to the well-being of the persons involved. Potentially, however, any unforseen event is dangerous and worthy of study. Moreover, detrimental psychological effects may result, even in the absence of physical damage. What seems to be called for is a broader conception of what constitutes an accident, and of what should be considered as its possible causes and effects.

Accident Reports

An analysis of available accident reports in the Norwegian Air Force revealed that a systematic study could hardly be made of them because of the great variations in quality and the failure to gather enough relevant data. Above all, data on the assessment of human factors in the performance of pilots and other personnel were lacking.

When investigating boards are appointed ad hoc, as they are in Norway, it would seem particularly important to regulate their functioning and advise on their work so as to obtain comprehensive and standardized accident reports. Such reports would provide a more adequate basis for systematic studies in this field. A tentative guide has been worked out, therefore, to facilitate the reporting as well as the logical and psychological analysis of predisposing and precipitating factors in aircraft accidents.

The guide is based on an attempt to identify the "logical universe" within which all possible causes of accidents can be grouped. Three domains were selected as the primary grouping: Materiel, Operation, and Command. Within each of these domains, sub-categories were then specified as follows:

I. Materiel

1. Manufacture

a. Inadequate design and directives
b. Inadequate execution of design and directives

2. Maintenance

a. Inadequate directives
b. Inadequate execution of directives

II. Operation

1. Experience

a. Inadequate training program
b. Special factors concerning practice

2. Predispositions or reaction tendencies

a. Persistent personality traits
b. Temporary reaction tendencies

III. Command

1. Giving of orders

a. Inadequate directives
b. Inadequate intermediation of directives

2. Execution of orders

a. Misunderstanding of orders
b. Neglect of orders

These sub-categories may, of course, be divided into more detailed groupings. It is of particular importance that the analyses always be carried back to human factors in the chain of causes. Defects in technical components are no exceptions to this rule. The same thing can be said about bad weather, which is sometimes mentioned as a cause of accidents. If a plane crashes because of bad weather, it is due to inadequate weather reporting or to inadequate interpretation of the reports (if there was no special order to go through with the mission regardless of weather conditions). In the first case, the meteorologist is in the spotlight, then, the operational domain for relevant questions. If the problem is one of interpretation the same kind of questions may be asked, but, this time, the pilot is the center of attention.

Such a guide may prove helpful to boards of investigation in ensuring proper attention to all the factors which may be relevant to aircraft accidents. It should contribute especially to improved coverage and understanding of the human factors involved in flying safety, and provide a better basis for judging when psychologists are consulted in connection with an accident. The guide has already proved to be helpful in working out a form for the recording of data.

Near-Accidents

The term "near-accident" has been chosen to distinguish an accident without resulting physical damage in the conventional sense. More specifically, a near-accident is said to have occurred when a pilot has experienced an unforeseen situation, with or without temporary loss of control, which under somewhat different circumstances could have led to an accident.

It may be assumed that near-accidents occur much more frequently than accidents leading to investigations and reports, and thus provide a much larger source of data. Moreover, the pilot always survives to tell the story. But, is he able or willing to tell? The first approach gave "no" as an answer. After thorough orientation and instruction, experienced pilots were asked to give written descriptions of their near-accidents. Only one out of fifteen turned in a report.

The next step was to have civilian psychologists interview a sample of 60 pilots who were informed that their reports would be treated as confidential. Under these conditions a total of 230 near-accidents was reported, the average being close to four per pilot.

Each psychologist interviewed about three pilots a day. At first, the subjects usually had nothing to report. Perhaps the chief factor conducive to communication was the realization that a near-accident need not be hyper-dramatic in order to be of interest. Yet, there is no lack of dramatic incidents in the material, nor of reports on immediate or delayed fear reactions in connection with the events.

The information given was not always detailed enough for a satisfactory analysis as to causes. However, if the confidence of the pilots can be obtained and preserved, this approach opens up a new and valuable source of information.

Within the domain of command, some near-accidents were ascribed to unsatisfactory or irresponsible disposition of leaders, others to insufficient briefing, and still others to a failure of support from towers, radio stations, landing control systems, etc.

Within the operational domain were found a lack of flying experience (either in general or on the specific type of aircraft in question), insufficient checking before take-off, inadequate observation or attention, failures of memory, perceptual disorientations, cognitive insufficiency, and poor maneuvering due to nervousness, panic, indisposition, etc.

It is interesting to note that those near-accidents which could be ascribed to materiel had already been reported and acted upon.

Data obtained from the analysis of near-accidents may be used in several ways. Aircrews can be informed about the types of incidents which have occurred, and how full-fledged accidents were avoided. Recommendations for training and for other preventive measures can be suggested. Last, but not least, many near-accident findings can be used as starting points for more specific research on flying safety.

Lateral Dominance

Leads for research may also come from individual contacts with pilots in connection with a general care-of-the-flyer program. Thus, among the "problem

cases" referred to the Psychological Division it was noticed that left-handedness occurred rather frequently; a check revealed that about one-third of the group was left-handed. The pilots' malfunctioning often seemed to have some connection with their left-handedness.

A survey of all pilots on active duty was then conducted by means of a 66-item questionnaire on lateral dominance. According to the criteria used, 7.6 per cent of this group was classified as left-dominant (as compared with 31.6 per cent of the referred problem cases). Statistically, the difference is significant at the .02 level of confidence, using the binomial test.

British aviation psychologists have stressed the fact that the pilot's task is to a large extent a perceptual one. It should be added that his job is also one of reacting to perceived signals (i.e., interpretation), and that the perceptual task involves auditory, spatial, and intellective-integrative as well as visual factors. A more detailed study of laterality would appear to have a bearing on all the variables mentioned.

The fact of being left-handed is not important in itself. The term merely signifies that a person uses his left hand more often or more readily than his right, and for many individuals this is the only peculiarity which can be observed. However, many left-handed subjects also show such symptoms as stuttering, word-blindness, awkward and uncertain movements, etc. On the other hand, many persons have such symptoms without being left-handed, although it seems to be commonly believed that left-handed individuals show them more often than others. Correspondingly, many right-handed pilots were found who have difficulties in flight similar to those of left-handed pilots, and many left-handed pilots who perform successfully in the aircraft.

Although left-handed pilots, as a group, seem to have more problems than others, some of them, nevertheless, adjust successfully. The question therefore becomes: in what ways do the successful left-handed pilots differ from the unsuccessful ones? And, further: do some of the unsuccessful right-handed pilots represent cases with problems of lateral dominance?

In the search for a fruitful point of view, concentrating attention on what may be called dysfunctions in directional and sequential behavior was found useful. Looking up and down, left and right, back and forth in a space exemplifies "directional" behavior. "Sequential" tasks involve the performance of actions which cannot be carried out simultaneously, e.g., speaking vs. listening to a speaker. The two types of behavior may be intermingled in practice, making analysis difficult.

Reversals in the conceptualization, organization, or performance of directional and sequential tasks seem to be of particular importance in the behavior of pilots in flight. Such reversals, here conceived as centrally determined tendencies to react in a manner opposite to that which is normal, can occur in the motor, visual, auditory, or ideational realms. These tendencies may lead either to doing the wrong thing at the right moment, or to a slowing down of reaction time, with the result that the right thing is done too late.

It may be presumed that learning and practice lead to more adequate performance, and thus tend to cover up any underlying predispositions which may exist. These may become manifest, however, in emergency situations or in the early stages of pilot training. The problem at hand is to find methods for the early diagnosis and measurement of such dysfunctions.

As a first approach to the study of lateral dominance in connection with vision, the method developed by Jasper, utilizing the phi-phenomenon, was employed. The retest reliability of this method proved unsatisfactory, but in a recent American study satisfactory reliability was obtained by improving fixation conditions. The same study showed that subjects classified as right- or left-eyed by the Jasper procedure are classified as cortically right- or left-dominated, respectively, by the improved method. This test is therefore considered to be a promising one for the purposes of the study.

Other approaches, such as the use of more comples perceptual and sensory-motor tests, are still in the blueprint stage. In order to bring out the dysfunctions more clearly, it may prove useful to administer some of the tests under conditions of low atmospheric pressure.

Summary

It has proved so difficult to increase the efficiency of existing pilot selection procedures that improvement of flying safety is better approached through the study of in-service performance and the analysis of accidents.

Examination of accident reports collected by the Norwegian Air Force revealed such great variations in the quality and scope of data recording that the systematic studies planned by the authors could not be carried out. There was a general failure to collect enough relevant information, particularly in the area of human factors. To facilitate the standardization and increase the comprehensiveness of accident reporting and analysis, the authors have developed a tentative guide for the classification of accidents.

The guide groups the various possible causes of accidents into three main categories: Materiel, Operation, and Command. Within each category are included more specific factors relating to (a) equipment manufacture and maintenance; (b) experience, training, and modes of reaction; and (c) the giving and execution of orders. It is held that the accident analyst must always probe deeply into human factors, which enter ultimately even into accidents attributable at first glance to such causes as bad weather or equipment failure.

The authors' use of near-accidents as a data source is described. Individual interviews with a sample of 60 pilots yielded reports of 230 near-accidents. Such reports contain much valuable information. The data obtained can be used in instruction and training, and as a starting point for further research.

Research on lateral dominance in pilots is also described. It was found that of the pilots referred to the Psychological Division as "problem cases," 31.7 per cent were left-handed. This percentage was significantly larger (.02 level of confidence) than that characterizing the general population of pilots (7.6 per cent). This finding suggests two questions: (a) In what ways do left-handed pilots who have adjusted satisfactorily differ from those who have not? (b) Do some unsuccessful right-handed pilots represent cases with problems of lateral dominance?

It is hypothesized that overt or latent dominance dysfunctions may cause reversals in the pilot's conceptualization, organization, or performance of directional or sequential behavior, and thus cause difficulties in flight. A beginning study of lateral dominance using visual tests is described, and plans for future research are outlined.

Résumé

Il est si difficile d'améliorer davantage les techniques courantes de sélection des pilotes qu'il est préférable d'aborder la question de la sécurité de vol sous l'angle de l'étude du comportement en service et de l'analyse des accidents.

A l'examen, les rapports d'accidents compilés par la Force Aérienne norvégienne manifestaient de telles divergences quant à la qualité et à l'étendue de l'enregistrement des données que les études systématizues projetées par les auteurs n'ont pu être menées à bien. La quantité d'informations pertinentes n'était pas suffisante, notamment en ce qui concerne les facteurs humains. Les auteurs, en vue de faciliter la normalisation et d'accroître le champ de la relation et de l'analyse des accidents, ont mis au point un premier essai de guide en matière de classification des accidents.

Ce guide groupe les diverses causes possibles d'accidents en trois catégories principales: le matériel, les opérations, et le commandement. Chaque catégorie comporte des facteurs plus spécifiques relevant de (a) la fabrication et l'entretien de l'équipement, (b) l'expérience, l'entraînement, et les modes de réaction, et (c) les ordres et leur exécution. On estime que l'analyste des accidents doit toujours sonder attentivement les facteurs humains, lesquels interviennent en fin de compte même dans les accidents imputables de prime abord à des causes telles que les intempéries ou les défaillances techniques.

Les auteurs expliquent comment ils ont recouru à des "accidents manqués" comme source d'information. Des entretiens individuels avec un échantillon comprenant 60 pilotes, ont mis en lumière 230 accidents manqués. Les rapports qui en ont été dressés contiennent des informations précieuses. Les données recueillies peuvent être utiles lors de l'instruction et de l'entraînement, et constitues un point de départ pour des nouvelles recherches.

Les recherches sur la dominance latérale parmi les pilotes sont également décrites. On a trouvé que 31.7 pour cent des pilotes renvoyés devant la Division Psychologique comme "cas-problèmes" étaient gauchers. Ce pourcentage est considérablement plus élevé (niveau de confiance 0,02) que eclui caractérisant l'ensemble de la population pilote (7,6 pour cent). Cette constatation soulève deux questions: (a) En quoi les pilotes gauchers qui se sont adaptés de façon satisfaisante diffèrent-ils de ceux qui n'ont pu le faire? (b) Certains des pilotes droitiers qui n'ont pas réussi constituent-ils des cas ayant des difficultés de dominance latérale?

On fait l'hypothèse qu'une anomalie de dominance avérée ou latente peut provoquer des inversions dans la conceptualisation, l'organisation ou l'exécution par le pilote du comportement directionnel ou séquentiel, et, partant, des difficultés de vol. Une étude préliminaire de la dominance latérale au moyen de tests visuels est décrite, et des projets de recherches futures sont esquissés.

Commentary by Discussant

Dr. Wilse B. Webb (U. S. A.):

In regard to the first part of this paper, concerning psychological categorization of the causes of aircraft accidents, let me offer a sincere "bon voyage" to my Norwegian colleagues as they set sail into these uncharted waters. But, let me also report some perils of the venture, as learned from first-hand experience with the navigational hazards of their route.

The needs expressed in the paper by Mr. Jarl and his colleagues have long been recognized. As early as the 1920's, the U. S. National Advisory Committee on Aeronautics introduced a system of coding accidents due to pilot errors which was used with varying degrees of consistency by the Navy, the Air Force, and the Civil Aeronautics Authority until about 1941. At that time, both the Navy and the Air Force substituted a descriptive system in which the actual behavior of the pilot was recorded without reference to underlying causes. Attempts to seek out the psychological factors responsible for accidents were not abandoned, however, and since the beginning of World War II about 20 studies using a variety of categorization techniques have been performed.

The perils of which I spoke, and to which so many of these studies succumbed, relate not to the types of categories used, but to their reliability. Obviously, if a classification system cannot be applied so as to yield consistent results, it is of limited use. The results of some 19 American studies can be briefly summarized by saying that the agreement among judges applying the same system of coding to the same accidents has been found to range between 20 per cent and 70 per cent. A primary factor appears to be the number of coding categories utilized: the larger the number of categories, the greater the amount of disagreement among the judges.

A study which I reported recently before a meeting of the Aero Medical Association represented, in regard to my own research in this area, what may be called a last dying gasp. The paper was called "Further Attempts at Coding Aircraft Accidents," and it was so entitled because it represented my sixteenth attack on this problem.

In this investigation, we attempted to maximize reliability of coding by (a) restricting ourselves to only one type of accident, (i.e., carrier landing accidents), (b) deriving our coding system from intensive interviews concerning such accidents, and (c) using only psychologically trained observers. With all of these precautions, the overall amount of agreement obtained among our three judges was approximately 30 per cent.

Those results were disappointing, but they were edifying. Personally, I have now concluded that success in achieving satisfactorily reliable coding of aircraft accidents is most unlikely. Although we can reproduce the accident itself, we cannot reproduce the pilot's state of mind at the time the accident occurs, and it is on the latter that human factors coding systems are based. Thus, while wishing Mr. Jarl and his colleagues the best of luck in their research, I should like to suggest that they carry out an early check on the reliability of their coding system.

In regard to the second section of their paper on the study of near-accidents, the authors are to be encouraged and applauded. In our experience, near-accidents

have been found to constitute a most valuable source of data. Again, however, a warning. Data based on a pilot's memory, which in the case of a near-accident may be affected on the one hand by pride and on the other by emotional factors, must be treated with all due regard for their fallibility.

On the last section of the paper, dealing with lateral dominance as an accident predictor, I should like to make one brief comment. This concerns the general question of whether aircraft accidents can be satisfactorily predicted from pilot-centered measures. In a recent review of this problem I concluded that extremely few accidents in operational flying are predictable from measures taken on individual pilots. By the time men have flown some 500 hours, the unfit among them have been rather thoroughly screened out by the original selection procedures, by failures in training, by self-elimination, and by accidents themselves. Of course, a very few men may pass all these hurdles while remaining basically inept. For the most part, I believe it is reasonable to assume that post-training accidents result not from any basic inability to handle the aircraft, but from situations to which no individual of a similar level of experience could respond, or from extremely transitory states such as fatigue, inattention, momentary sets, moods, and the like. If this is true, then it seems to me that the prediction of aircraft accidents from relatively remote measures such as lateral dominance would not have too great a chance of success.

In closing, I must apologize for the generally negative nature of my remarks. They do not result from a general pessimism, as many of my colleagues well know, nor from any proven inappropriateness of the procedures suggested in the previous paper. Rather, they represent the voice of a man somewhat battered by ten years of experience in a tremendously complex field.

General Discussion

Prof. Torsten Husen (Sweden):

In connection with the remarks made by the authors regarding lateral dominance, I should like to refer to an investigation carried out in Sweden by Prof. Arne Trankell. In this study it was shown rather definitively, I think, that lateral dominance is not related to measured personality traits. Although a large and varied battery of personality tests was employed, no significant correlations with lateral dominance were obtained. It was also shown, for a population consisting of all children of a certain age in the city of Stockholm, that left-handedness is not related to stuttering, word-blindness, or other deviations often mentioned in the literature as being correlated with this variable. Finally, Trankell presented evidence indicating that lateral dominance is inherited in a relatively simple way. I should personally be inclined to go along with Dr. Webb and question the importance of left-handedness per se as a cause of accidents in problem cases of the type studied by the authors.

Dr. H. K. Knoepfel (Switzerland):

As a psychiatrist for the Swiss Air Force, I have had an opportunity to interview a number of pilots who have had near-accidents, and to test them with such projective instruments as the Rorschach and the TAT. Although my sample is too small to permit a statistical statement, it seemed to me that many of these pilots tended to have "accident-prone" personalities. That is to say, their interpersonal relations were disturbed, they had a higher-than-average tendency to get into legal difficulties in civilian life, and they often had histories of near or small accidents while driving, while engaging in sports, and so on. I did not, conversely, find it possible to predict the occurrence of accidents from knowledge of the personality. When accident-prone personalities of the type described were encountered, however, I made it a practice to inquire into the history of previous accidents. If a summation of small accidents was found, we eliminated the individual or did not accept him for training.

Reply

Mr. V. Coucheron-Jarl:

With respect to the question of lateral dominance, I should like to point out that we have not yet progressed to the point of attempting to use laterality as a predictor of accidents. We have merely run across the fact that of those pilots referred to us by their fellow officers as "problem cases," one-third turned out to be left-handed. We are just beginning to be interested in this problem and to try to find out something about it. Some left-handed pilots are poor performers in the aircraft while others adjust satisfactorily. We should like to discover why this difference exists. Only when we know more about the reasons for the phenomenon will it be feasible to consider possible applications of the findings.

AUDITORY AND VISUAL SEARCH PROBLEMS

Donald Wallis
Department of the Senior Psychologist, The Admiralty
United Kingdom

There are few military operations where the active phase of attack or defense is not preceded by a relatively long period of just looking for relevant targets. Generally, the search is conducted at ranges beyond those at which the unaided human senses are effective, so that radar, sonar, and other engineered systems of location must be employed. Yet, it is still necessary to use a human operator to detect the symbolic information displayed by such equipment. In a wide variety of military situations, therefore, the following tasks are required of operators, possibly for several hours at a stretch: (a) discrimination of characteristic auditory or visual signals at the lowest possible signal-to-noise ratios, and (b) maintenance of continual alertness to detect the occasional and unpredictable occurrence of such signals.

Search activities of this kind have become known as "vigilance tasks" (Mackworth, 1950). Ideally, a vigilant operator would always detect targets at signal strengths corresponding to his psycho-physical threshold of response for the partirular kind of signal displayed. The published results from many studies and operational trials suggest, however, that this is rarely, if ever, so (Mackworth, 1957). On the other hand, it is not possible to say at all precisely what the trend of performance will actually be in a new search or vigilance task. The variety of conditions under which experiments have been carried out precludes any but the broadest of generalizations. There seems to be little doubt that a significant deterioration in search efficiency will occur as time-on-watch increases. The particular aspects of performance which will suffer most, and after what lengths of time, can be determined only empirically for any specific operational task.

The example of one series of experiments illustrates some of the experimental problems and methods which have been under investigation in the Royal Navy. The results are then reviewed briefly in the light of auditory and visual behavior in general.

Description of Experiments

The operational circumstances leading to the study were broadly as follows. Two operators were airborne for several hours in the separate cockpits of an aircraft. Each was required to maintain an intensive watch on two independent search systems. Thus, each man was faced with the double vigilance task of scanning a radar plan-position-indicator (P.P.I.) display, and of simultaneously listening through headphones for occasional pure-tone signals from other equipment. Both visual and auditory signals were noise-masked.

Although complete realism could not be provided in a laboratory study, efforts were made to simulate effectively those features of the situation which expe-

rience had shown systematically influence performance. A full-scale mock-up of the aircraft cockpit was constructed, and was equipped with apparatus which provided a reasonably realistic presentation of both the visual and auditory signals (Plate 1). The simulated radar display presented small echoes of uniform size against a constantly changing background of "noise" and "clutter." Appropriate bright-up and decay characteristics were associated with the time-base rotation rate of 43 rpm. The auditory signals consisted of three 20-msec pulses at 835 c/s, separated by 1 1/2-sec intervals and appearing against a constant "white noise" background at 75 db above .0002 dynes/sq.cm. Ambient temperature, noise, and brightness were controlled in the cockpit at values characteristic of a normal "twilight" patrol.

The experimental design and performance criteria were selected with the following hypotheses in mind: (a) that the maintenance of a vigilant watch would be assisted by interrupting the visual search (a passive, perceptual task) with interpolated activity of an active, intellectual nature, and (b) that vigilance would be decreased by the operator's lack of expectancy that a relevant signal was always imminent.

Naval aircrew—12 officers ("observers") and 12 ratings ("telegraphists"—acted as subjects. A first group of ten men experienced two 3-hour experimental sessions during which hypothesis a was tested. Later, the same subjects underwent five successive 3-hour sessions, each separated by a one week interval. The second, comparable group of men was given five 3-hour search tests at intervals of one day only. All the test periods, and each succeeding hour of any one period, were strictly comparable, except that no man experienced exactly the same pattern of relevant stimuli more than once.

Possible sources of unwanted variance due to practice effects, spatial location of the visual (P.P.I.) signals, apparatus irregularities, etc., were controlled in

Plate 1.--Subject in Cockpit, Showing Radar Visor and Plotting Materials.

the design. Table 1 shows how the visual and auditory signals were deployed: ten signals of each kind were presented at random time intervals during every hour of a test. No two stimuli ever occurred simultaneously.

TABLE 1
PLAN OF EACH SUCCESSIVE HOUR DURING A 3-HOUR SEARCH

Serial Minute Number	Nature of Tasks	Number of Signals
0-5	Auditory Search + Navigation	1 auditory
6-20	Auditory + Visual Search	2 auditory, 4 visual
21-30	Auditory Search + Navigation	2 auditory
31-60	Auditory + Visual Search	5 auditory, 6 visual

It will be noted that while the auditory search was maintained continuously for three hours, there was a total of 15 minutes hourly when the visual watch was replaced by navigational exercises; these the subject took up and discarded when instructed by the experimenter through an intercom channel. However, in order to test hypothesis a, ten subjects were given two 3-hour sessions of which one was conducted without introducing any break at all in either the visual or auditory search.

Various measurements and observations were recorded to throw light on the performance trends. Visual search efficiency was measured first as the number of times a signal was repeated, i.e., successive echo "paints", before being seen. Auditory efficiency was expressed as the signal-to-noise ratio (decibels) at which signals were just heard. Here, the measurement technique consisted of raising the auditory signal level every four seconds in 1-db steps from an intensity well below the threshold of response. This is a modified method of limits as used first in studies of vigilance by E. Elliott (1957) and by P. Bakan (1955).

Noise-masked psychophysical thresholds for the two kinds of signals used in the search runs were measured for each subject in the cockpit immediately before and after every session. These thresholds were designated "expectancy" scores, since the subjects were naturally warned and alerted for the occurrence of signals during the taking of all such measurements. Scores obtained under the search or vigilance conditions were termed "non-expectancy," since here the subjects did not know just when a signal was going to occur. By comparing the two corresponding sets if scored for any subject on any trial, it was possible to observe by how much the "effective threshold" (cf. Elliott, 1957) of response to any visual or auditory signal had risen above the psychophysical threshold.

Supplementary data were collected throughout the experiments by direct observation of the subjects (unknown to them), by an interview after each run, and by a special technique employed to record the amount of gross physical movement of the trunk.

Summary of Results

The main quantitative results were related to several different criteria of search efficiency. The statistical analyses were based on the sets of "expectancy" and "non-expectancy" visual and auditory scores computed for each hour of each experimental session. Visual scores were taken as the medians of responses to ten signals, with the best responses being indicated by the smallest numbers. Auditory scores consisted of the means of responses to ten signals. Here, the most negative figures (signal-to-noise ratios) indicate the most effective performance.

First, general performance trends were analyzed hour by hour throughout each 3-hour session, and between daily or weekly trials. The results showed that under the conditions of these experiments, no progressive decrement in performance occured on either the visual or the auditory task (see Fig. 1 and Fig. 2). This was true whether average response level or within-hour variability provided the index. There were also no consistent differences between the first and second half-hours of a trial.

There was one significant exception to this pattern, however. When the navigational exercises were eliminated, and the operators had to scan visually (as well as listen) continuously for three hours, they responded significantly less well on

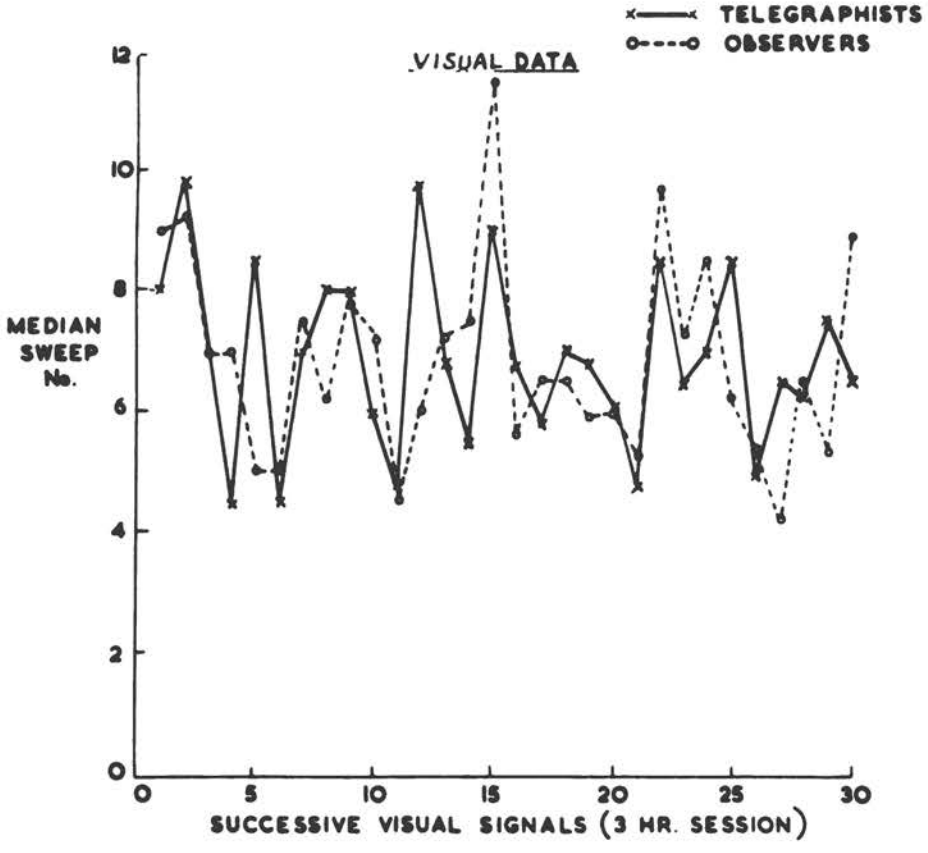


Fig. 1.--Responses to Successive Signal Presentations (Daily Sessions)

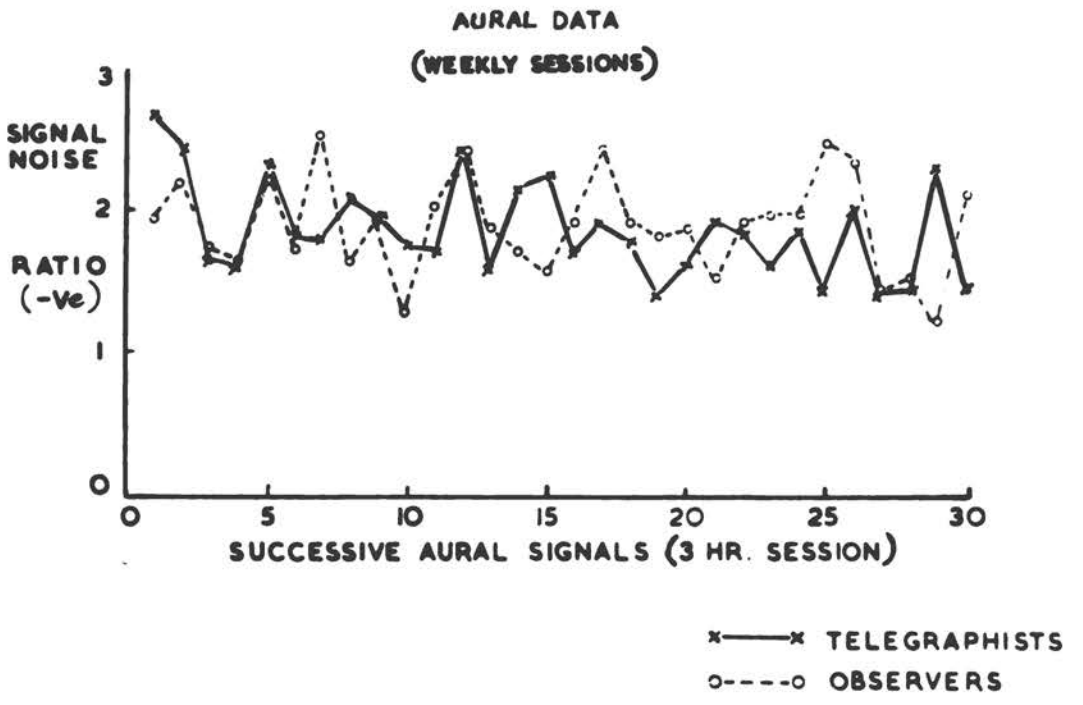


Fig. 2

both tasks all the time. Moreover, their performance tended to deteriorate as time went on (see Table 2).

Generally, all subjects were highly self-consistent in their response levels on any one occasion, although they were inclined to differ from one occasion to another. Auditory detection usually took place at a signal-to-noise ratio within ± 2 db of the mean for three hours. Yet, it was clear from individual records that brief but serious lapses of vigilance sometimes occurred in all subjects.

A further check was carried out to determine whether visual and auditory effective thresholds were "positive," i.e., whether the non-expectancy scores for each subject always tended to exceed his corresponding expectancy scores. This proved to be the case on both tasks. As soon as a search began, the subjects responded less effectively than when they were fully alerted and warned to expect a signal, as during the psychophysical tests before and after each trial (Fig. 3).

Individual differences were marked in both search tasks. It was notable, however, that corresponding visual and auditory expectancy and non-expectancy scores were significantly correlated (tau = + .67 and + .70, respectively). Apparently, the individual variability under search conditions was due as much to pre-existing differences in detection ability as to differences in degree of alertness. However, by comparing the orders of merit for individual efficiency at the two tasks, it was established that one could not predict an individual's relative efficiency at one task from his relative efficiency at the other.

The visual responses were also analyzed separately in relation to the spatial position of the displayed signals. The result has direct implications for tasks such as radar scanning, where the location as well as the timing of a signal is unpredictable. As might be anticipated, detection was least efficient in the periphery of

594869 O - 61 - 3

TABLE 2

ANALYSIS OF DATA FROM PRELIMINARY EXPERIMENT

Visual Data: Analysis of Variance of Mean Hourly Performance (Log Scores)

Source of Variation		Sum of Squares	df	Mean Square	F	P
Continuous vs. Discontinuous Radar	(CD)	.028558	1	.028558	4.507	<.05
Observers vs. Telegraphists	(OT)	.181610	1	.181610	28,659	<.01
1st vs. 2nd vs. 3rd Hours	(H)	.131525	2	.065763	10.378	<.01
CD x OT		.001315	1	.001315		
CD x H		.016297	2	.008149	1.286	n.s.
OT x H		.014369	2	.007185	1.134	n.s.
CD x OT x H		.006089	2	.003045		
Total Between Groups		.379763	11			
Between Individuals		.804127	9	.089347	14.099	<.01
Residual		.247124	39	.006337		
TOTAL		1.431014	59			

Aural Data: Analysis of Variance of Mean Hourly Performance (S/N ratios)

Source of Variation		Sum of Squares	df	Mean Square	F	P
Continuous vs. Discontinuous Radar	(CD)	2.280	1	2.280	7.780	<.01
Observers vs. Telegraphists	(OT)	2.057	1	2.057	7.020	<.05
1st vs. 2nd vs. 3rd Hours	(H)	2.110	2	1.055	3.601	<.05
Pooled Interactions		2.647	7	.378	1.290	n.s.
Total Between Groups		9.080	11			
Between Individuals		56.190	9	6.243	26.121	<.01
Residual		11.430	39	.293		
TOTAL		76.700	59			

Significance of Between-Hours Differences

Average Visual performance differs significantly between any two hourly periods, with the second hour the worst ($t_{19} > 22$, $p < .001$)

Average Aural performance is significantly worse during the 3rd hour than during the second. ($t_{19} = 4.89$, $p < .001$); the first and second hours do not differ significantly.

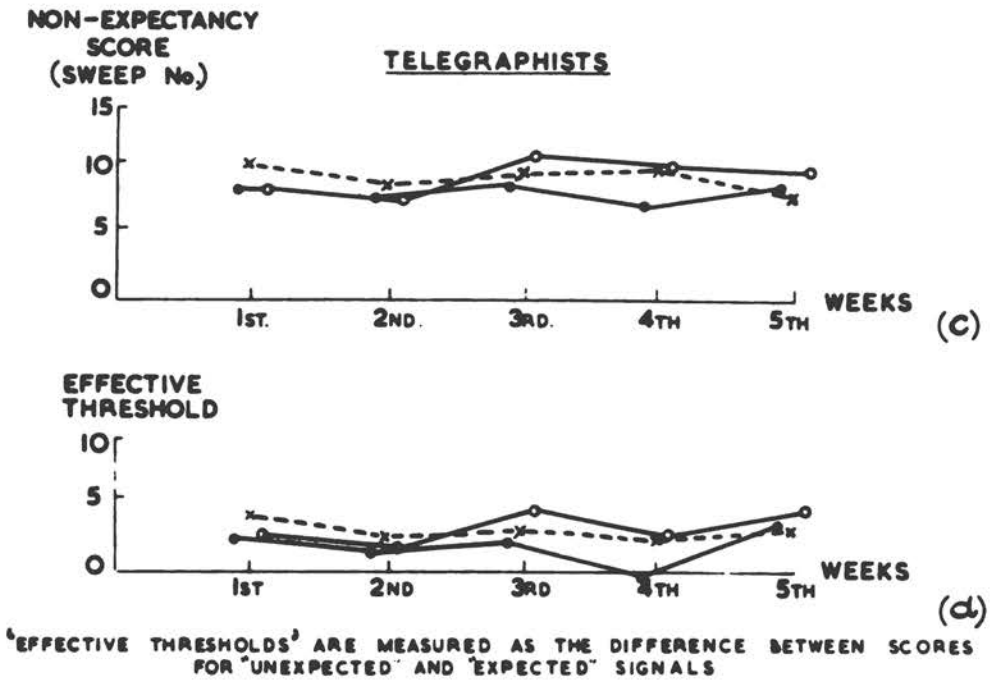
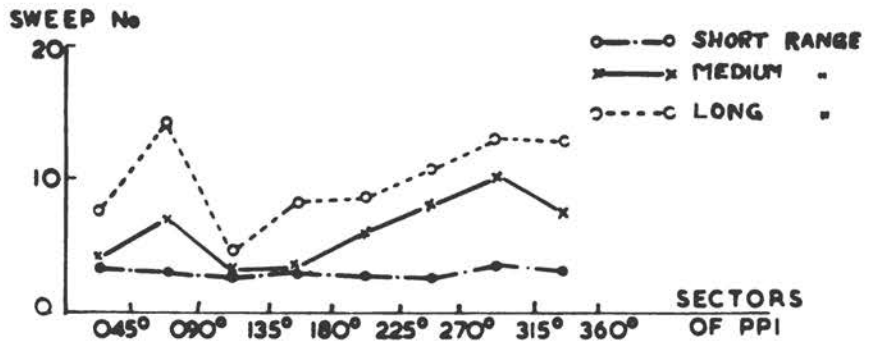


Fig. 3.--Comparison Between 'Non-Expectancy' Scores and 'Effective Thresholds'

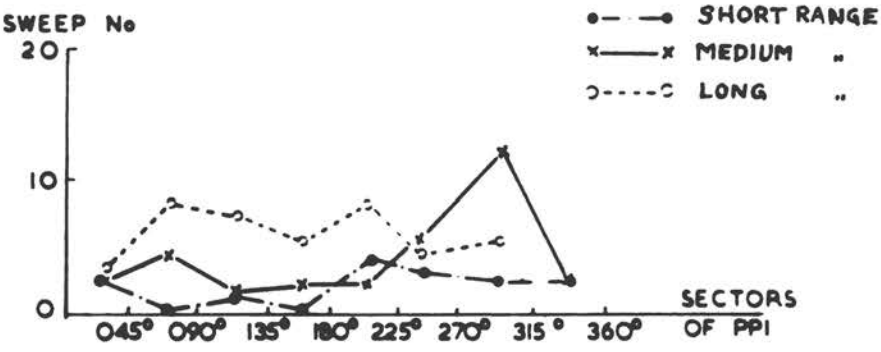
Visual Data: Weekly Sessions

the display, where a greater area had to be searched. Comparison of expectancy and non-expectancy scores showed that this disadvantage is accentuated under prolonged search conditions (Fig. 4). This is a form of decrement not revealed by the former analyses.

Finally, it can be mentioned that one way of compensating for the recurrent lapses of alertness which occur on both the visual and auditory tasks is to have two men doing the job at the same time. The cumulative percentage frequency curves presented in Fig. 5 show that the probability of detection is approximately doubled (improvement being particularly marked for stimulus intensities near the threshold), if two operators carry out the task simultaneously. The reason for this, of course, is that the attentional lapses of different individuals are not temporally correlated. While operators do tend to "go to sleep" at unpredictable intervals, fortunately they do not all tend to do so at the same time.



MEDIAN EXPECTANCY SCORES (ALL SUBJECTS)



EFFECTIVE THRESHOLD SCORES (ALL SUBJECTS)

Fig. 4.--Visual Detection in Various Display Sectors and Range Bands
(Data from Weekly Sessions)

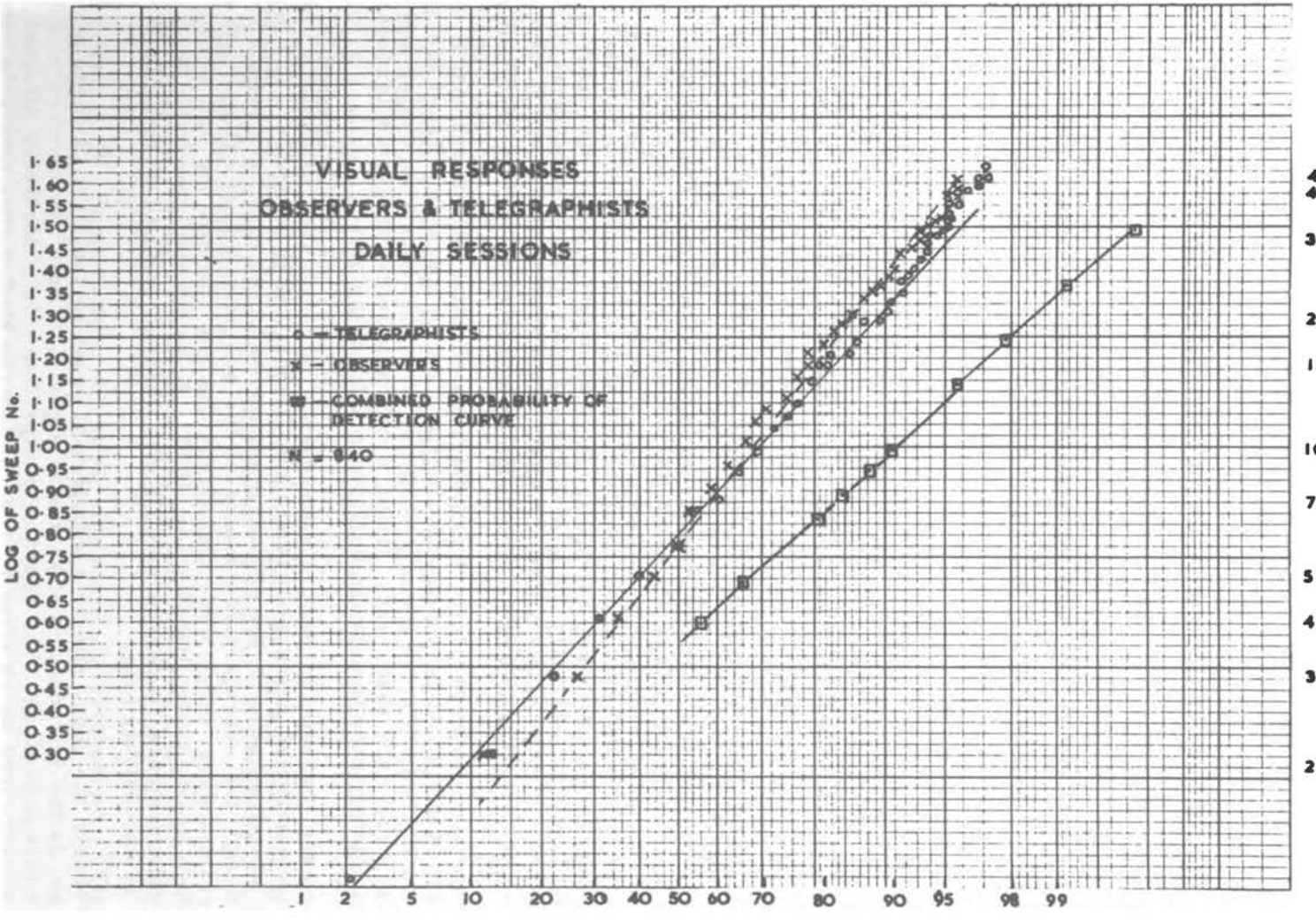


Fig. 5.--Cumulative Percentage Frequency of Given Responses.

Discussion

These experiments were conducted under circumstances which have not previously been brought together in studies of search activity. First, the task and environmental conditions more closely resembled a practical situation than is usually possible or even desirable in the laboratory. Second, the subjects were relevant service personnel and highly-motivated. Third, two perceptual tasks were carried out simultaneously. There are reasonable grounds for supposing that the first two factors favoured good performance; reasons are given below for supposing that the third also did.

There are clearly two broad factors affecting search efficiency. The first is the specific perceptual skill involved. Even when fresh and fully-alerted, operators differ in detection ability to an extent which would make it profitable either (a) to seek methods of selecting the more able personnel, or (b) to devise special training to improve "threshold" responses. The latter approach is likely in practice to be administratively preferable.

Equally important, however, is the factor of maintenance of alertness. Here, an eclectic viewpoint is necessary in relation to search tasks generally, in the light of the present results and the various viewpoints which have been expressed in the vigilance literature (Adams, 1956; Berlyne, 1951; Bexton, Heron, & Scott, 1954; Broadbent, 1953; Deese, 1955; Elliott, 1957; Mackworth, 1957).

Uninterrupted visual search for long periods should be avoided. All the signs of visual fatigue, boredom, sleepiness, and frustration were present in the appearance of and reports from the subjects tested in this condition. The interpolation of a different, non-vigilance task for brief interludes was welcomed, and was enough to remove any progressive deterioration (cf. Mackworth, 1950). Auditory search, however, seems to be much more resistant, and does not suffer in the same way from continued application (cf. Elliott, 1957). The subjects responded equally well when simultaneously searching visually or pursuing navigational exercises; a change of supplementary task appeared not to upset the auditory skill, which was soon claimed to be "almost automatic," and to demand no conscious effort to attend.

There seem to be two important influences affecting all search tasks. First is the marked reduction in expectancy imposed on an operator. From his viewpoint, there is always a much lower probability of signals occurring during a search (non-expectancy) task than, say, during a psychophysical threshold (expectancy) test. This factor tends to depress performance from the outset.

Secondly, perhaps in part, a consequence of diminished expectancy, is the tendency toward brief, more or less random fluctuations in attention. The occurrence of this phenomenon in the present experiments has already been noted. Moreover, although the average performance of the subjects did not suffer, even after five successive days, there were significant changes towards the end of this period in their hourly variability (Fraser, 1952). It seems highly probable that short though serious lapses in search activity will increase even in an auditory task, if the vigil is excessively long.

Probably the most effective counter, at least to the first of these adverse tendencies, is the introduction of synthetic signals. There are dangers and difficulties in this procedure. The operator's expectancy is undoubtedly affected both by the frequency with which signals occur, and by the kind of time intervals between

them (Baker, 1956; Deese, 1955). The subjects, for example, probably reacted to a combined signal rate of 20 per hour, rather than to separate visual and auditory rates of 10 per hour. There is evidence indicating that a rate of 20 signals per hour leads to better performance than do lower rates (Bowen & Woodhead, 1956). These may have been contributory reasons for the comparative success of the subjects. More experiments are certainly needed, however, if a full understanding of this mechanism is to be attained.

References

Adams, J. A. Vigilance in the detection of low-intensity visual stimuli. J. exp. Psychol., 1956, 52, 204-208.

Bakan, P. Discrimination decrement as a function of time in a prolonged vigil. J. exp. Psychol., 1055, 50, 387-390.

Baker, C. H. Biasing attention to visual displays during a vigilance task: a summary report. Admiralty: Royal Naval Personnel Res. Comm. Rep. No. 876, October, 1956.

Berlyne, D. E. Attention, perception and behavior theory. Psychol. Rev., 1951, 58, 137-145.

Bexton, W. H., Heron, W., & Scott, T. H. Effects of decreased variation in the sensory environment. Canad. J. Psychol., 1954, 8, 70-76.

Bowen, H. M., & Woodhead, M. M. The effect of signal rate on target detectability. Air Ministry: Flying Personnel Res. Comm. Rep. No. 955, January, 1956.

Broadbent, D. Classical conditioning and human watch-keeping. Psychol. Rev., 1953, 60, 331-339.

Deese, J. Some problems in the theory of vigilance. Psychol. Rev., 1955, 62, 359-368.

Elliott, E. Auditory vigilance tasks. In symposium: "Vigilance." Advancement of Science, 1957, 13, 393-399.

Fraser, D. C. A study of deterioration of performance in aircrew. M. R. C. Applied Psychology Research Unit Report No. 185. Unpublished Manuscript, Cambridge, England, 1952.

Mackworth, N. H. Researches on the measurement of human performance. M. R. C. Special Report Series, No. 268. London: H. M. Stationery Office, 1950.

Mackworth, N. H. Some factors affecting vigilance. In symposium: "Vigilance." Advancement of Science, 1957, 13, 389-393.

Summary

The research dealt with the study of behavior in a situation requiring prolonged visual and auditory vigilance. Its ultimate purpose was to develop methods for improving performance in military tasks, such as radar and sonar operation.

In the experiments described, 24 naval aircrew—12 officers ("observers") and 12 ratings ("telegraphists")—served as subjects. They were tested in a scale-model cockpit under noise, temperature, and brightness conditions characteristic of a normal "twilight" patrol. The cockpit was equipped with a simulated radar scope and headphones which mediated, respectively, the visual and auditory search tasks (Plate 1). During the experimental sessions the subjects were required simultaneously to scan the radar and listen for the auditory signals. Both the stimulus displays were noise-masked.

All experimental sessions were three hours in duration. During each hour of a session, the visual and auditory signals were presented at random time intervals according to the plan shown in Table 1. No two signals ever occurred simultaneously.

One group of ten subjects underwent five sessions separated by one-week intervals. The second, comparable group was given five tests separated by one-day intervals. During each hour of each session, the subjects were required periodically to interrupt their visual search to carry out a total of 15 minutes' work on navigational exercises. To test the hypothesis that interpolation of an active intellectual task of this nature helps to maintain vigilance, the group of ten subjects also underwent two experimental sessions, during one of which no navigational work was required.

To determine whether the maintenance of alertness is affected by the operator's relative state of "expectancy," noise-masked psychophisical thresholds for the visual and auditory signals were measured for each subject in the cockpit immediately before and after each experimental session. These "expectancy" scores (obtained under conditions in which the subjects were alerted before the presentation of each signal) were compared with the "non-expectancy" scores of the search runs (during which the signals occurred at unpredictable intervals) to provide an indication of whether the subjects' "effective thresholds" of visual and auditory response exceeded their psychophysical thresholds.

Analysis of within-hour (first vs. second half), between-hour, and between-session performance trends showed that, in general, no progressive decrement occurred on either the visual or the auditory task (Fig. 1 & Fig. 2). When the navigational exercises were eliminated, however, and the subjects had to watch (as well as listen) continuously for three hours, performance was consistently poorer on both tasks, and tended to deteriorate with time (Table 2).

It was further found that performance under expectancy conditions was always superior to performance in the non-expectancy or search situation (Fig. 3).

Individual differences were marked in both search tasks. Corresponding sets of visual and auditory expectancy and non-expectancy scores were significantly correlated (tau = + .67 and + .70, respectively), indicating that performance is affected by pre-existing differences in detection ability as well as by differences in degree or alertness. The subjects' efficiency on one task could not, however, be predicted from knowledge of their efficiency on the other.

An analysis of the effect of stimulus location on visual search performance showed that detection was least efficient for signals appearing on the periphery of the simulated radar display. Comparison of expectancy and non-expectancy scores showed that the difference between center and periphery tended to increase with time (Fig. 4).

Brief but serious lapses of attention occurred periodically in all subjects. Since the lapses of different individuals are not temporally correlated, it is possible to compensate by having two operators carry out the vigilance task simultaneously. This greatly increases the probability of signal detection, particularly, at stimulus intensities near the threshold (Fig. 5).

The unusually good performance of the subjects can be attributed to the following factors: (a) the experimental situation closely simulated operational conditions; (b) the subjects were highly motivated; (c) two vigilance tasks were carried out simultaneously; (d) active, intellectual tasks were interpolated into the visual watch; (e) the rate of signal presentation was relatively high (20 per hour).

To maximize search efficiency in practical situations it is suggested that operators be selected on the basis of pre-existing differences in detection ability, or that special training be given to improve threshold responses. It is also suggested that prolonged periods of uninterrupted visual search be avoided, since performance tends to deteriorate under these conditions. Auditory vigilance appears to be much more resistant to the effects of continued application.

Two influences affecting search tasks in general are pointed out: (a) lowered expectancy on the part of the operator under random as opposed to predictable stimulus conditions, and (b) periodic fluctuations of attention. It is considered probable that the frequency of attentional lapses will increase, even for auditory tasks, if vigils are excessively long. Since stimulus frequency appears to affect expectancy levels, it is suggested that vigilance can be improved by introducing synthetic signals which serve to increase the rate of stimulus presentation to approximately 20 per hour.

Résumé

Les recherches ont porté sur l'étude du comportement dans une situation exigeant une vigilance visuelle et auditive soutenue. Elles ont pour objet en dernière analyse la mise au point de méthodes d'amélioration du rendement de certaines tâches militaires, telles que l'opération du radar et du sonar.

Les expériences décrites ont pris pour sujets 24 membres d'équipage aéronaval: 12 officiers ("observateurs") et 12 conscrits (télégraphistes"). Les sujets ont été soumis à des tests en carlingue expérimentale dans des conditions de bruit, de température et d'éclairement caractéristiques d'un vol de patrouille "crépusculaire" normal. La cabine était équipée d'un faux écran de radar et de casques d'écoute transmettant les tâches respectivement visuelles et auditives à accomplir (Planche 1). Pendant les expériences les sujets étaient invités à surveiller le radar et à être simultanément a l'écoute des signaux auditifs. Les deux séries de stimuli étaient brouillées par du "bruit."

Chacune des expériences durait trois heures. Au cours de chacune de ces heures, les signaux visuels et auditifs étaient émis à des intervalles arbitraires selon le programme repris au Table 1. Deux signaux ne se présentaient jamais simultanément.

Un groups de dix sujets a fait l'objet de cinq expériences se succédant à une semaine d'intervalle. Le second groupe, comparable, a fait l'objet de cinq expériences se succédant à un jour d'intervalle. Au cours de chaque heure d'expérience, les sujets étaient invités à interrompre périodiquement leur exploration visuelle pour se consacrer à des exercices de navigation de 15 minutes au total. Au fins de vérification de l'hypothèse selon laquelle l'interpolation d'un travail intellectuel actif de cet ordre aide à maintenir la vigilance, le groupe de dix sujets a également été soumis à deux séances expérimentales au cours desquelles aucum travail de navigation n'était imposé.

Pour déterminer si le maintien des facultés d'alerte est affecté par l'état relatif d'attente de l'opérateur, on a mesuré pour chacun des sujets en cabine, immédiatement avant et après chaque expérience, les seuils psychophysiques des signaux visuels et auditifs brouillés par du "bruit." Ces résultats "d'expectative" (obtenus dans des conditions de mise en alerte des sujets avant chaque émission de signal) ont été comparés avec les résultats "d'inexpectative" des séances expérimentales (pendant lesquelles les signaux étaient émis à intervalles non prévisibles) de façon a fournir une indication sur le point de savoir si les "seuils effectifs" de réponse visuelle et auditive des sujets excédaient leurs seuils psychophysiques.

Des analyses intrahoraires (comparant les deux demi-heures), interhoraires et interséances des tendances du comportement ont fait ressortir qu'aucune décroissance progressive ne marquait en général, soit la tâche visuelle soit la tâche auditive (Fig. 1 et Fig. 2). Toutefois, après élimination des exercices de navigation, et les sujets ayant dû maintenir leur attention visuelle (et auditive) continûment pendant trois heures, leur comportement était systématiquement plus pauvre pour les deux tâches et tendait à s'affaiblir avec le temps (Table 2).

Il a de plus été constaté que le comportement en état d'expectative était toujours supérieur au comportement en état d'inexpectative ou d'exploration (Fig. 3).

Les différences individuelles étaient accusées pour les deux tâches. Les cotes "expectatives" et "inexpectatives" étaient en corrélation significative (tau = .67 pour la tâche visuelle et .70 pour la tâche auditive), ce qui indique que le comportement est affecté par des différences pré-existantes dans le facultés de détection aussi bien que dans le degré d'état d'alerte. Le rendement des sujets pour une des tâches ne pouvait, cependant, se laisser prédire à partir de la connaissance de leur rendement pour l'autre.

Une analyse de l'influence de la localisation du stimulus sur le comportement dans l'exploration visuelle a révélé que la détection était moins efficace pour des signaux apparaissant à la périphérie de l'écran de radar. Une comparaison entre les résultats expectative et inexpectative fait apparaître que l'écart entre centre et périphérie tendait à s'accroître avec le temps (Fig. 4).

Périodiquement tous les sujets ont éprouvé des chutes d'attention brèves mais sérieuses. Ces chutes étaient sans corrélation temporelle entre les individus, une compensation est possible si deux opérateurs s'acquittent simultanément de la mission de vigilance. Cette formule accroît sensiblement la probabilité de la détection du signal, notamment à des intensités de stimulus voisines du seuil (Fig. 5).

Le comportement exceptionnellement bon des sujets pourrait être dû aux facteurs suivants: (a) les conditions expérimentales étaient étroitement proches des conditions réelles; (b) les sujets étaient fortement "motivés;" (c) deux missions de vigilance devaient être accomplies simultanément; (d) des tâches intellectuelles actives étaient introduites dans le courant de la veille visuelle; (e) la fréquence des signaux était relativement élevée (20 par heure).

Pour optimaliser l'efficacité de l'exploration dans la réalité, on suggère que les opérateurs soient sélectionnés en fonction de différences pré-existantes dans les facultés de détection, ou qu'un entraînement spécial soit conçu pour améliorer les réponses au niveau du seuil. On suggère également d'éviter toute période trop prolongée d'exploration visuelle continue, le comportement tendant à s'affaiblir dans ces conditions. La vigilance auditive apparaît beaucoup plus résistante aux effets d'une application soutenue.

Deux influences affectant en général les missiones de vigilance sont mises en lumière: (a) une expectative réduite de la part de l'opérateur dans des conditions de stimulation arbitraires par opposition à des conditions prévisibles, et (b) des fluctuations périodiques de l'attention. On estime probable que la fréquence des chutes d'attention s'accroisse même pour des missions auditives si les veilles sont excessivement prolongées. La fréquence de la stimulation paraissant affecter le niveau d'expectative, il est suggéré que la vigilance pourrait être améliorée par l'introduction de signaux synthétiques destinés à porter la fréquence de la stimulation jusqu'à un taux d'environ 20 par heure.

Commentary by Discussant

Dr. Paul Bakan (U. S. A.):

I should like to begin by commenting on Mr. Wallis' distinction between "expectancy" and "non-expectancy" scores, which is the difference between the usual type of psychophysical measurement and thresholds taken in the course of a vigil, where the subject does not know just when a signal will occur. This difference points up an important aspect of vigilance performance, namely, that in the vigilance situation the subject is more or less free to determine his own behavior. He can listen or not listen, watch or not watch, and it is this possibility of freedom, usually absent in the typical measurement of a psychophysical threshold, which may account for the difference between expectancy and non-expectancy scores. In the usual psychophysical experiment, the subject is alerted before each stimulus presentation, and it is very unlikely that he will not look or listen when the experimenter says, in effect, "Here comes a stimulus." Lapses of alertness would appear to be much more probable in the vigilance situation, where the individual does not know exactly when a signal will come.

With respect to the results concerning the lack of a performance decrement over time, I think that what Mr. Wallis has shown regarding the effects of the navigational task can be compared to the findings of Mackworth and others on the effect of rest periods. The navigational exercises constitute a self-paced task which requires the subject to do something very different from what he does while watching or listening for signals.

In this same connection, I should like to mention a further possible reason for the absence of performance decrements in the visual-plus-auditory stimulus situation. As Mr. Wallis has pointed out, it may be that performance did not fall off here because in the double-search situation the total number of signals presented was 20 per hour. We know from other research that with signal rates of this order the level of performance remains relatively constant. It might now, incidentally, be of interest to go further and to compare visual-plus-auditory vigilance at a signal rate of 20 per hour, with vigilance for visual and auditory signals presented alone at the rate of 20 per hour. Such a comparison would throw light on the question of whether two-modality vigilance is better or worse than one-modality vigilance with signal rate held constant. I know of no data on this question and think it is one of considerable interest.

Since signal frequency is an important determiner of alertness, Mr. Wallis has suggested the possibility of using synthetic signals to improve or maintain vigilance. This plan would no doubt work, but it has its difficulties, especially in practical military situations. For example, increasing the number of signals by the addition of "false-alarm" pips in an air defense radar system could result in a confusion which might prove prohibitively expensive.

An alternative to the use of false signals is the introduction of an irrelevant secondary task which, though clearly different from the main task, nevertheless involves an increase in the total number of effective stimuli emanating from the display. I have carried out an experiment along these lines with quite successful results. The primary task was one of listening for sequences of three odd numbers in a random digit sequence, and the secondary task was that of pushing a button at every occurrence of the single digit "six." The additional requirement of listening for "sixes" did not add to the difficulty of the primary task, but it increased the signal frequency about ten-fold. Under these conditions there was a significant increase in the number of primary signals detected, indicating that the introduction of a relatively simple secondary task which increases the total number of effective stimuli results in improved performance on a more difficult primary task.

Now, an interesting thing about this result was its relation to individual differences in personality. In the above experiment, I fortunately had available for each subject a measure of introversion-extraversion. Analysis of the data in relation to this variable showed that the subjects who benefited most from the introduction of the secondary task were those who tended to score high on extraversion. The differential response of extraverts and introverts can be understood by referring to the difference in performance of these two groups in the absence of the secondary task. This difference was most marked in the first half-hour of the task, with the extraverts missing more signals during this early period. Since the extraverts started at a lower level of performance, they did not show the marked decremental trend which characterized the introverts. The effect of the secondary task on the performance of the extraverts was most marked during the first half-hour, and most of their overall gain with the secondary task was due to improvement in this part of the vigil. The secondary task produced no change in the performance of the introverts during the first half-hour, probably because performance was already at a high level, thus making further improvement unlikely.

It should be noted, incidentally, that in the absence of the secondary task the group with the greater decrement in performance—the introverts—was nevertheless the group with the better overall performance. This means that decrement is not necessarily a good criterion of effectiveness in a vigilance task if the aim is to maximize overall efficiency.

In an attempt to rationalize the above findings, I should like to offer an account of what I think may be happening in the vigilance situation. First, I believe it is necessary to distinguish between "listening" and "hearing," and between "looking" and "seeing." It is quite possible for one to be aware that stimuli are present without, as it were, absorbing their meaning. For example, it is quite possible for you to be aware that I am talking without knowing what I am saying. This distinction can be considered as a function of what we may very broadly call "attention." When we say that an individual is "attending" to a stimulus display, we usually mean that he is making some sort of discrimination among the stimuli within that display. If these attentional responses, like overt motor responses, are liable to reinforcement and extinction, then it may be that vigilance decrements occur at low signal frequencies because the subject's attentional responses become extinguished due to a lack of sufficient reinforcement. If we conceive of the signals as reinforcing stimuli, it is apparent that, when the number of signals is low, the subject's watching or listening responses are only infrequently reinforced. Consequently, the responses may tend to extinguish.

When an individual stops attending to a display he tends to become sleepy. Drowsiness, which is very characteristic of subjects in vigilance experiments, is accompanied by a de-differentiation of the stimulus display. This de-differentiation would become absolute if the subject actually fell asleep. However, the subjects usually do not fall asleep, probably because they are socially motivated to stay awake and follow the instructions of the experimenter. To stay awake, I believe that they try to re-differentiate their environment by paying attention to other stimuli—stimuli which are irrelevant to the main vigilance task and which tend to distract attention from it. The origin of these irrelevant stimuli can be either external or internal: things in the environment which are not in the display, or imaginal activity, such as day-dreaming or mind-wandering. Either type of distracting stimulus may cause decrements in vigilance performance when there is a low rate of signal emission from the display.

Thus, the relationship of introversion-extraversion to performance on vigilance tasks may be a function of a differential susceptibility to extinction on the part of introverts and extraverts. The attentional responses of extraverts may tend to extinguish sooner, so that these subjects stop attending to the display and start paying attention to irrelevant stimuli earlier in the vigil.

General Discussion

Dr. Norman H. Mackworth (U. K.):

Some recent work by Dr. C. H. Baker of Toronto has shown that vigilance is affected not only by the number of signals presented, but also by their spacing in time. It appears that with total time held constant irregularly spaced signals result in poorer performance. For example, if there is a 20-minute period containing ten signals, and then a 10-minute gap without signals, followed by another 20-minute period containing ten signals, and so on, performance is definitely poorer than it is for signals which are more evenly spaced throughout the period of the watch. So, it appears that the temporal spacing of the stimuli must also be given consideration in the study of vigilance.

Another factor which should be mentioned is that of task alternation. We know from previous research that the introduction of rest periods helps to counteract the decrements which occur with prolonged work on tasks involving fine visual discrimination. Interestingly enough, some recent work by Dr. E. L. Saldanha shows that decrements can also be reduced by alternating one fine visual discrimination task with another of similar type. Thus, subjects who alternately carry out two visual discrimination tasks perform better, on the average, than those who work at the same task throughout the experimental session.

Lastly, I should like to comment on Mr. Wallis' most interesting finding of a greater decrement for the peripheral portions of the radar display. It seems to me that three factors might be operating here. The first is the narrowing of the effective field of vision which has been found to occur under such conditions as that of oxygen lack and high ambient temperature. If, for example, one has prolonged work plus high atmospheric temperature, then it would seem particularly probable that signals on the periphery would be missed. The second possible factor is the one mentioned by Dr. Bakan, i.e., that the operators look at the signals without seeing them. This phenomenon has often been observed in eye-movement studies. Thirdly, it is possible that the subjects simply do not look at the edges of the display as often as they look at its center. I should like to suggest that this finding of a greater peripheral decrement be followed up with a more detailed analysis aimed at determining the reasons for its occurrence.

Dr. Joseph Zubin (U. S. A.):

Some recent work carried out by my colleagues and myself would appear to be relevant to the problems under discussion. We were interested in finding out whether the notoriously slow reaction time of schizophrenics is solely a function of motivation, or whether differences in refractory periods are also involved. We therefore exposed matched groups of normals and schizophrenics to an alternating series of visual and auditory stimuli equated for intensity across modalities. The problem was to determine whether the response latencies of schizophrenics differed from those of normals: (a) when a stimulus of one modality was followed by another stimulus of the same modality, and (b) when stimulus modalities were alternated. It was found that when only three seconds were allowed between stimuli there was no significant difference in reaction time between normals and schizophrenics. However, when the stimuli were separated by 7-second intervals, the schizophrenic subjects showed a significantly longer reaction time following a change of modality. These results suggest that each individual may have a characteristic refractory period for two-modality stimulation within which he experiences interference from previous stimuli, but beyond which he shows no interference. It is possible that this factor is related to individual differences in vigilance performance.

Reply

Mr. Donald Wallis:

With reference to Dr. Bakan's comments about synthetic signals, it should be pointed out that the synthetic signals used in our experiments were quite clearly synthetic to the operator. In other words, they were sufficiently different from the real signals to prevent the possibility of confusion. Our "synthetic" signals, therefore, correspond in principle to the "secondary" signals employed by Dr. Bakan in his own research.

Dr. Bakan also suggested that two-modality vigilance be compared with one-modality vigilance with the rate of stimulus presentation held constant. We have not attacked this problem within a single experiment, but we do have evidence, from a stparate study of conar operators, that a signal rate of 20 per hour is sufficient to prevent decrements in performance on an auditory task alone. Furthermore, I have recently carried out some work on radar operation which shows that a signal rate of just less than 20 per hour also prevents deterioration on a visual task alone.

With respect to the comments of Dr. Mackworth, I would agree that the range of attention, or effective field of vision, tends to narrow down as one becomes less vigilant, and that this mechanism may have contributed to the observed differences in threshold for the periphery as compared to the center of the radar display. I also think that Dr. Bakan's explanation in terms of operant conditioning is a sound one, and that a tendency toward extinction of attentional responses to the periphery may have occurred as the experimental sessions progressed.

THE EFFECTS OF STRESS ON LOWER NEURON ACTIVITY[1]

R. Margaria, T. Gualtierotti, and D. Spinelli
Laboratory of Physiology, University of Milan
Italy

A major problem in investigating stress is that the phenomenon is so difficult to define. Moreover, symptoms of stress can be evoked by different causes and manifested through various mechanisms. Such symptoms can result, for example, from physical exercise, heat, cold, anoxia, lack of sleep, lack of food, and various emotional reactions such as dissatisfaction and anxiety. Thus, a fruitful approach to the problem of stress and a common method of study are not easy to find.

A specific physiological alteration or group of alterations common to all types of stress has not yet been discovered and perhaps does not exist. However, an objective impairment of function and the particular subjective feeling commonly referred to as "fatigue," appear to accompany all stress states. It is plausible to hypothesize that these phenomena indicate the occurrence of a characteristic change or changes in the activity of the central nervous system. On this hypothesis, a series of investigations were undertaken which aimed at detecting any such alterations which may occur under various stressful conditions.

The approach was through the study of spinal reflexes because these responses are relatively easy to measure in man and exemplify the most elementary as well as the most fundamental activity of the central nervous system. The functioning of the C. N. S. can, in fact, be reduced to the passage of impulses from neuron to neuron through more or less complicated neural circuits. Spinal activity is particularly interesting from this point of view because all centers of the nervous system ultimately impinge on the efferent pathway and on its last neurons—the motoneurons of the spinal cord. Furthermore, recent research suggests that human monosynaptic spinal reflexes reflect the state of the C. N. S. as a whole. In several studies (Languth, Teasdall, & Magladery, 1952; Magladery, Teasdall, Park, & Languth, 1952; Teasdall, Park, Languth, & Magladery, 1952; Teasdall, Languth, & Magladery, 1952), relationships have been observed between changes in spinal reflexes and injuries at different levels of the spinal cord, brain stem, and telencephalon.

Procedure

In the present investigation, monosynaptic reflexes were produced in the sciatica region by direct stimulation of a mixed spinal nerve. Such stimulation gives rise to two temporally distinct responses measurable as action potentials in the muscle. The first response is due to excitation of the motor fibers, and the second, a delayed, reflex reaction, is due to excitation of the sensory fibers. If an ex-

[1]This research was supported under Contract AF 61(514)-637 by the Office of Scientific Research of the United States Air Force Air Research and Development Command.

594869 O - 61 - 4

perimental stimulus is repeated at two points a known distance apart along a spinal nerve, the muscle action potentials resulting from each stimulus will occur at different times, thus permitting exact determination of conduction speed in the sensory and motor fibers. The other two components of total reflex time—muscular end-plate delay and central or spinal delay—can then be calculated by appropriate substitution and subtraction of values.

The experimental procedure consisted of stimulating a mixed spinal nerve, both peripherally and centrally, at locations a precisely determined distance apart, and computing the end-plate delay, the central reflex time, and the speed of conduction in the sensory and motor fibers according to the method outlined above. Figure 1 schematizes the procedure utilized. A very brief, high-voltage, square-wave electrical impulse was applied to the spinal nerve through electrodes at P and D, and the resulting action potentials were measured by means of electrodes applied to the muscle at M. In the graphs at the right of the figure, Pd and Dd equal, respectively, the times between application of the stimulus (St) at P and D, and direct contraction of the muscle. Pr and Dr represent the total reflex times for stimuli applied at P and D respectively.

The stimulus applied to the nerve had to be very brief because the central delay in the monosynaptic spinal reflex is so short (one to two milliseconds) that large errors of measurement would have been introduced if the experimental stimulus had not been of a briefer duration. An impulse, therefore, of 75 microseconds (representing about 5 per cent of the central synaptic time) was utilized. The intensity of the stimulation also had to be carefully regulated, since too strong a stimulus, leading to excitation of all the efferent fibers, would have blocked is virtually

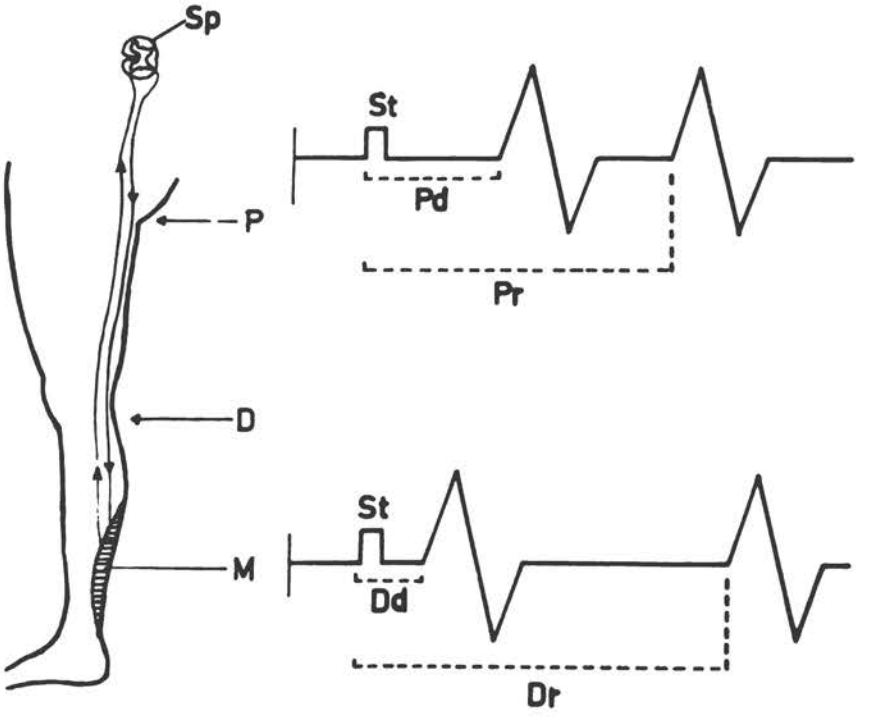


Fig. 1.--Method Used for Measuring Various Temporal Components of the Human Monosynaptic Spinal Reflex

complete; after about 75 milliseconds its excitability is roughly halfway towards recovery. Thus, to study the Renshaw suppression effect two stimuli, separated by an interval of about 75 milliseconds, were presented to the motoneuron. The amplitude of the second reflex response was then measured and compared with the amplitude of the first response under normal conditions and under various stress conditions of the experiments.

The action of the Renshaw suppression effect represents a very important aspect of the functioning of the central nervous system. As shown in previous research, it is because of this mechanism that the spinal motoneurons are unable to fire, in either voluntary or reflex muscular contraction, at frequencies higher than 12 to 15 impulses per second. This limited discharge frequency also characterizes many other types of cell in the nervous system, including those of the cortex. It was for these reasons, and because the phenomenon is known to be very sensitive to changes in experimental conditions, that Renshaw suppression was included among the variables studied.

The stress conditions imposed in the present research comprised hypoxia, hypoglycemia, muscular fatigue due to prolonged physical effort, various drugs (caffeine, bartiburates, alcohol), and lack of sleep.

Results

Hypoxia

The subjects were required to breathe a 6 per cent oxygen mixture for about 18 minutes, observations being taken every five minutes from the beginning of the experiment up to 20 minutes after the return to normal air breathing. It was found that hypoxia acted to increase sensory and motor conduction time and to decrease end-plate delay and net central reflex time (see Fig. 2). In a few cases, motor conduction speed and end-plate delay tended to increase, but this was later discovered to be due to the hypocapnia, which sometimes supervenes during hypoxia. If hypocapnia does not occur, efferent conduction speed and end-plate delay always decrease under hypoxia.

The effect of the above changes on total reflex time is determined by algebraic summation of the various time components. It can be noted that the decreases in end-plate and spinal delay are not completely cancelled by the increases in sensory and motor conduction time, so that hypoxia tends to produce an overall decrease in total reflex time.

Figure 3 presents data on the Renshaw suppression effect. In recording this effect, the amplitude of the second reflex response produced by the experimental stimulation (bottom line) was always expressed as a percentage of the first reflex response produced (top line). The three sets of actual tracings in the upper part of Fig. 3 were taken before, after ten minutes of hypoxia, and after a return to normal atmospheric conditions. The tracings show that Renshaw suppression is considerably diminished by hypoxia, this being indicated by the fact that the measured amplitude of the second reflex response is appreciably larger during hypoxia than it is during the prior breathing of ordinary air. As might be expected, the observed reduction of inhigition is only transient: measurements taken subsequent to the period of hypoxia show that suppression of the second response once again sets in when normal oxygen conditions are restored.

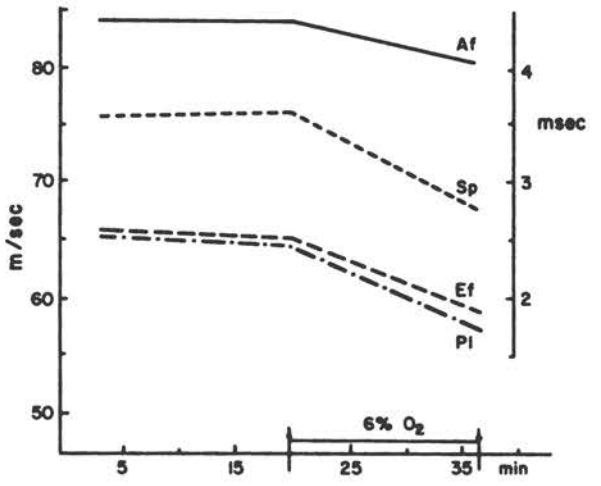


Fig. 2.--Effects of Hypoxia on Various Reflex Components

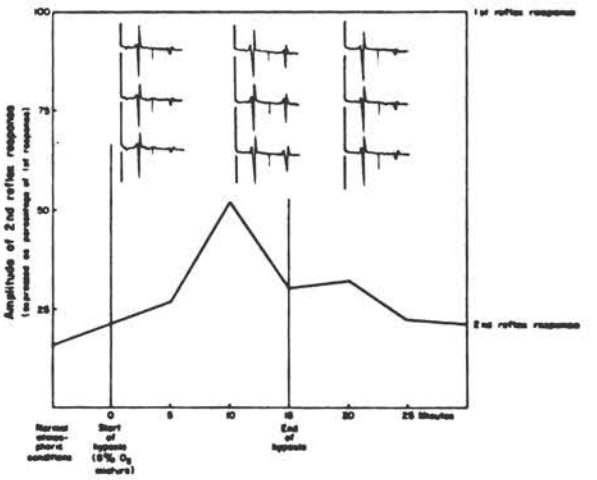


Fig. 3.--Effects of Hypoxia on Renshaw Suppression

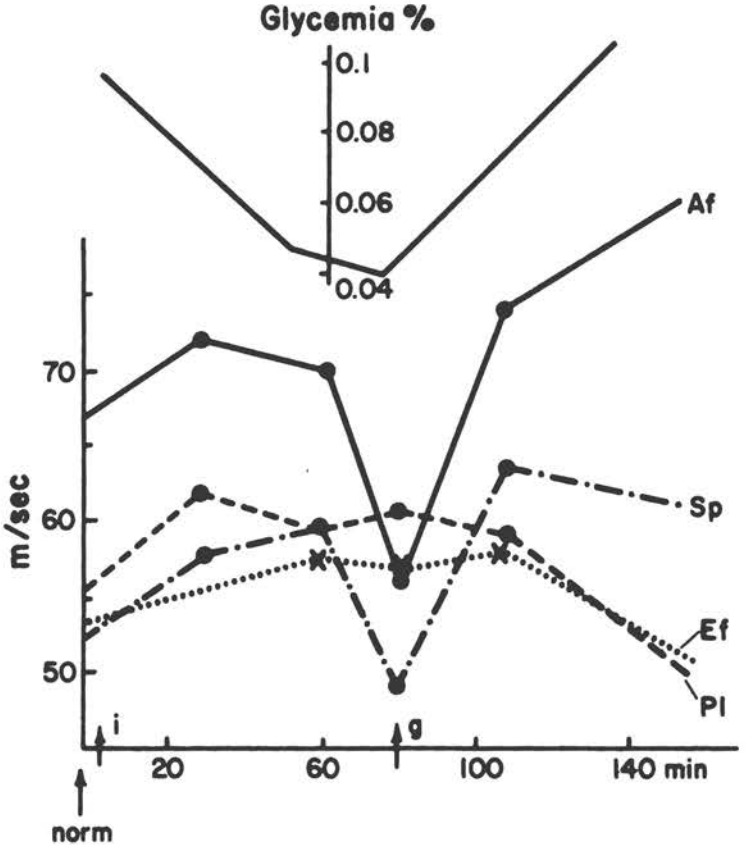


Fig. 4.--Effects of Hypoglycemia on Various Reflex Components

Hypoglycemia

A moderate degree of hypoglycemia (not exceeding 0.07 per cent), effected by insulin injection, produces an increase in motor and sensory conduction speed, and in the central and peripheral delays. Since the two kinds of change are opposite in direction, the total reflex time is not appreciably altered. Figure 4 illustrates the changes observed in the above variables after insulin injection (at i), and later glucose injection (at g).

Monosynaptic reflexes appear to be very sensitive to changes in glycemia. Even the small variations in blood sugar level which occur during the course of the day are accompanied by changes in spinal delay and in conduction speed along the nerves.

When the blood sugar level is more severely lowered, motor conduction speed and end-plate delay increase still further, but sensory conduction speed and central delay show an appreciable decrease, the lowest value of the latter being reached at the threshold of convulsions. Following the return to normal blycemia levels, after-effects can be observed for as long as 24 hours. It has been hypothesized that this prolonged after-effect is due to the action of supra-spinal centers, since the phenomenon is not observed in spinalized animals subjected to the same conditions.

Renshaw suppression is also highly sensitive to changes of blood sugar level, being greatly diminished by hypoglycemia.

Muscular fatigue

Exercise on a cycloergometer at a rate of 2.5 to 8 Kcal/min was performed for one to one and a half hours. Immediately after the exercise, and up to about an hour later, sensory conduction speed was increased by about 12 ± 0.5 percent. Motor conduction speed showed an opposite and approximately equal change, so that the two tended to cancel each other, and overall conduction time in the reflex pathways remained virtually constant. End-plate and central delays tended to decrease after exercise, although in a few cases the central time increased. Renshaw suppression was not studied under this experimental condition.

Effects of drugs

1. Caffeine. The subjects were given 250 mg of caffeine intravenously. Sensory and motor conduction speed did not change at any time after administration, and end-plate delay showed only a slight increase. Spinal delay was markedly reduced, the effect being found to last for several hours. This decrease was progressive for 15 minutes after injection, the new value then remaining steady for about four hours, after which there was a gradual return to normal. Along with the decrease in spinal delay, an increase in the amplitude of the reflex response was observed.

A severe reduction of Renshaw inhibition was also evident during the initial period of observation. Fifteen minutes after injection, the amplitude of the second reflex response was nearly 75 per cent of the first. After 80 minutes, however, the second reflex dropped to 20-25 per cent of the first response and remained at this level for several hours.

Caffeine is known to produce two characteristic effects: (a) deep polarization of the spinal cord with marked catelectrotonus, and (b) synchronization of synaptic

potentials. The second phenomenon may be responsible for the observed decrease of spinal delay in the subjects. Catelectrotonus acts to facilitate both excitatory and inhibitory synapsis at all junctions of the spinal cord; it would appear that after the transient block of the Renshaw inhibitory effect, this inhibition prevails.

2. Barbiturates. The barbiturate Diogenal was given intravenously in doses sufficient to synchronize the EEG at the frequency of the alpha rhythm. In such a state of (relatively light) narcosis, the observed reflex response is very similar to that characteristic of spinalized animals. Response amplitude is increased and remains highly constant. With larger doses the response tends to become smaller and more irregular.

Under the conditions of the experiment, barbituric narcosis produced no change in sensory or motor conduction speed. End-plate delay was observed to increase very slightly. Spinal delay and Renshaw inhibition increased more markedly. Such changes probably result more from supra-spinal action of the drug, e.g., elimination of cortical influences on the cord, than from any direct action on the spinal centers themselves. The increase of inhibition may be due to increased catelectrotonus, as under caffeine.

3. Alcohol. The subjects were given 125 cc of 95 per cent alcohol, with sufficient dilution, in the briefest time possible. Measurements of the effects on spinal reflexes were taken every 15 minutes after absorption. Throughout the period of the experiment the subjects showed signs of deep intoxication.

The main changes observed consisted of a small increase in motor conduction speed and of large increases in spinal delay and total reflex time. Virtually no change occurred in sensory conduction speed or end-plate delay. The increases in motor conduction speed and in central delay reached their maximum within the first half hour, after which the values remained constant until the progressive return to normal. The duration of the changes varied markedly from subject to subject, lasting from a minimum of two up to many hours after administration. Renshaw suppression was not studied.

Lack of sleep

A group of normal subjects, untrained to sleep loss, was kept awake for an entire night, no stimulating drugs or increased smoking being allowed. Analysis of the effects on the dependent variables revealed no change in neural conduction speeds or in end-plate delay. Central spinal delay, however, showed a remarkable increase of up to three times the average. (See Fig. 5). Total reflex time was, accordingly, also increased. A full night's sleep was necessary to bring about a restoration of spinal delay to normal values; brief periods of sleep did not suffice.

Under these conditions of sleep loss, Renshaw inhibition was severely reduced, the second reflex response manifesting an amplitude of about 78 per cent of the first. This effect also disappeared after a suitable period of sleep.

At present, no completely satisfactory explanation of the above results can be offered. The lengthening of spinal delay may be due less to increased transmission time at the individual synapses than to a general desynchronization of synaptic action at the spinal junctions. The latter phenomenon has been observed in animal experiments, in which it has been demonstrated that synchronization of spinal synapsis is facilitated by the action of higher neural centers in the brain (Gualtierotti, 1953).

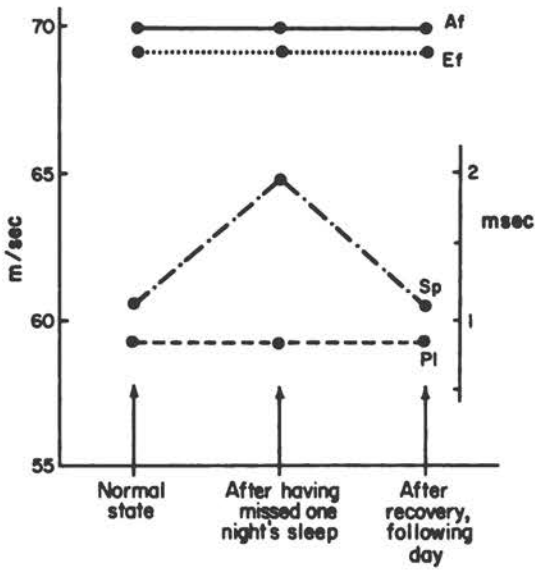


Fig. 5.--Effects of Sleep Loss on Various Reflex Components

With respect to the marked decrease of Renshaw inhibition under sleep loss, it can be hypothesized that the effect is due to increased catelectrotonus. However, this explanation requires that a likeness be postulated between spinal cord catelectrotonus and the action of the Renshaw cell. The results of the caffeine experiments provide some basis for such an assumption.

Discussion

The stress-inducing conditions studied in these experiments can be grouped into two categories with respect to their effects on the dependent variables:

1. Conditions which produced effects on the peripheral as well as on the central mechanisms involved in the reflex response;

2. Conditions which influenced central functioning alone, leaving the peripheral mechanisms unaffected.

Hypoxia, hypoglycemia, alcohol, and muscular fatigue fall into the first category. The second includes caffeine, barbiturates, and lack of sleep.

The (peripheral) changes in sensory and motor conduction speed observed under the conditions in the first group could be due either to direct effects exerted on the nerve fibers themselves, or to indirect effects emanating from changes in the functioning of the spinal cells. Because sensory conduction speed was calculated in the experiments from reflex responses provoked in circuits which included

spinal sensory centers, it is not possible, on the basis of the present data, to decide between these two explanations with respect to afferent transmissions. That direct action on the nerves is possible, however, was shown in the research by the fact that changes in motor conduction speed were observed. Here, no spinal mechanisms were involved in the determination of conduction time, so that the observed changes can be attributed only to an alteration in the functioning of the motor fibers themselves. Supporting this conclusion is the finding by Monnier and Lavigne (1953) that CO_2 and lack of oxygen can directly affect conduction in the nerve. Also relevant to the results for hypoxia is research indicating that anoxemia depresses the prejunctional structures and thus decreases the speed of transmission from nerve to muscle (Negrete, 1956). The increase in end-plate delay found under hypoxia in the experiments may be explainable in terms of this mechanism.

With respect to the peripheral effects of hypoglycemia, it can be noted that variations in nervous conduction due to changes in the amount of glucose present in the perfusing liquid have been clearly demonstrated in ganglia. In the absence of glucose, transmission is rapidly blocked in acutely denervated ganglia, especially if the ganglion is stimulated (Larrabee, Garcia, Edwards, & Horowicz, 1953), and a failure of conduction in the C fibers takes place (Horowicz & Larrabee, 1954). Insulin may also affect end-plate and neural conduction through the fall in plasma potassium which it entails (Dury, 1952).

The central changes observed under all the stress-inducing conditions could be due to purely loc l effects, to supra-spinal influences, or to both. In regard to supra-spinal influences, it is likely that the increase in spinal delay observed after sleep loss is attributable, at least partly, to effects produced on the sleep centers in the brain stem. These centers are known to be connected with reticular and hypothalamic centers which are capable of influencing the activity of the spinal motor neurons.

Insulin probably also exerts an effect on supra-spinal structures. Even when hypoglycemia is avoided, cortical electrical activity becomes slower and of greater amplitude after insulin injection, this effect being due to excitation of sub-cortical structures (Grenell & Wolbarsht, 1956). Hypoglycemia also acts to stimulate the hypothalamic centers and causes an increase in the adrenalin content of the blood, which directly activates the ascending and descending reticular system and the epimamillary region. This could result in a downward influence on the alpha and gamma neurons of the spinal cord, and, thus, on the muscle fibers and intrafusal system. On the other hand, experiments performed on spinalized cats have shown that hypoglycemia also acts at the spinal level. The effects of both insulin and glycemis are thus seen to be very complex, and it is not surprising that experimental manipulation of these variables produced impressive changes in the reflex response.

Turning to a consideration of local effects, it can be hypothesized that the observed changes in spinal delay were due to either (a) reduced speed of transmission at individual synapses, or (b) synchronization of all the synapses activated. The effect of insulin on spinal delay is probably explainable in terms of the first of these mechanisms, since insulin also produces delay at the end-plate, which can be considered as a form of synapse. Central delay due to caffeine is probably attributable to the second mechanism, since this substance is known to produce strychnine-like synchronization in the spinal cord.

Alcohol and barbiturates may have a double effect, producing changes of synchronization in the spinal centers directly, and electrotonus in the cord indirectly

by way of supra-spinal structures. Supporting the latter hypothesis are findings indicating that bartiburates act directly on the reticular formation and the brain stem sleep centers of Hess (Akert, Koella, & Hess, 1952).

In conclusion, it can be noted that during the initial stages, the central and peripheral changes produced under certain kinds of stress, e.g., hypoglycemia, tend to be mutually compensatory, so that the overall reflex time remains unaltered. At later stages, however, this compensation breaks down, and changes in total reflex time are observed. Under some conditions, such as hypoxia, hypoglycemia, and alcoholic intoxication, alterations in components of the reflex appear very early, and thus would seem to provide a sensitive index of stress. In order to interpret properly the present results in terms of organismic functioning, more data are needed on the effects produced on monosynaptic reflexes by other types of stress-inducing conditions, including psychological factors and various types of hormones.

References

Akert, K., Koella, W. P., & Hess, R., Jr. Sleep produced by electrical stimulation of the thalamus. Am. J. Physiol., 1952, 168, 260-267.

Dury, A. Effect of intravenous glucose on tissue water and ionic content and moderation of potassium plethora with epinephrine and insulin pretreatment. Am. J. Physiol., 1952, 171, 630-635.

Grenell, R. G., & Wolbarsht, M. Proc. XXth Int. Congr. Physiol., 1956, p. 369. (Abstract).

Gualtierotti, T. Synchronization of action potentials in the spinal frog. J. Physiol., 1953, 121, 106-116.

Horowicz, P., & Larrabee, M. G. Fed. Proc., 1954, 13, p. 72. (Abstract).

Languth, H. W., Teasdall, R. D., & Magladery, J. W. Electrophysiological studies of reflex activity in patients with lesions of the nervous system. III. Motoneurone excitability following afferent nerve volleys in patients with rostrally adjacent spinal cord damage. Bull. Johns Hopkins Hosp., 1952, 91, 257-266.

Larrabee, M. G., Garcia, J. R., Edwards, C., & Horowicz, P. Proc. XIXth Int. Congr. Physiol., 1953, 543-544. (Abstract).

Magladery, J. W., Teasdall, R. D., Park, A. M., & Languth, H. W. Electrophysiological studies of reflex activity in patients with lesions of the nervous system. I. A comparison of spinal motoneurone excitability following afferent nerve volleys in normal persons and patients with upper motor neurone lesions. Bull. Johns Hopkins Hosp., 1952, 91, 219-244.

Monnier, A. M., & Lavigne, S. Les fluctuations de l'excitabilité du nerf periphérique isolé et ses principaux facteurs. J. Physiol., Paris, 1953, 45, 198-199.

Negrete, J. M. Proc. XXth Int. Congr. Physiol., 1956, 674-675. (Abstract).

Teasdall, R. D., Park, A. M., Languth, H. W., & Magladery, J. W. Electrophysiological studies of reflex activity in patients with lesions of the nervous system. II. Disclosure of normally suppressed monosynaptic reflex discharge of spinal motoneurones by lesions of lower brain-stem and spinal cord. Bull. Johns Hopkins Hosp., 1952, 91, 245-256.

Teasdall, R. D., Languth, H. W., & Magladery, J. W. Electrophysiological studies of reflex activity in patients with lesions of the nervous system. IV. A note on the tendon jerk. Bull. Johns Hopkins Hosp., 1952, 91, 267-275.

Summary

Stress is difficult to investigate because its definition is elusive and its symptoms can be evoked by various causes and manifested through diverse mechanisms. Common to all stress states, however, are: (a) an objective impairment of function, and (b) subjectively experienced fatigue. This suggests the occurrence during stress of characteristic changes in the central nervous system. The present studies were aimed at detecting any such changes which may occur under various stressful conditions.

The approach was through the study of monosynaptic spinal reflexes, since these fundamental phenomena are easy to measure in man and appear to reflect the state of the central nervous system as a whole.

The experimental procedure involved the evocation of spinal reflexes in the sciatic region through direct stimulation, centrally and peripherally at locations a precisely determined distance apart, of a mixed spinal nerve (Fig. 1). Total reflex time and the intervals between stimulation and the resulting action potentials in the muscle of the leg were measured and substituted in appropriate equations to give indices of the following reflex components: (a) speed of conduction in the sensory fibers, (b) speed of conduction in the motor fibers, (c) muscular end-plate delay, and (d) spinal delay.

To stimulate the nerve, a high-voltage, 75-microsecond electrical impulse was applied by means of a specially constructed apparatus which permitted precise control of stimulus intensity and duration. Errors in the measurement of end-plate delay were avoided by use of a standardized method of localizing the electrodes placed on the muscle.

Also studied as a dependent variable was the Renshaw suppression effect—the brief inhibition of a motor neuron which results from its stimulation, immediately subsequent to firing, by an internuncial Renshaw cell. Effects of stress on Renshaw suppression were analyzed by comparing the relative amplitudes of the first and second reflex responses provoked by two stimuli applied to the motor neuron at an interval of 75 milliseconds (the time taken by the neuron to recover, after the first stimulation, approximately half of its normal excitability).

The stress conditions imposed in the experiments comprised: hypoxia, hypoglycemia, muscular fatigue, drugs (caffeine, barbiturates, alcohol), and lack of sleep.

The experimental results can be summarized as follows:

Hypoxia (produced by breathing a 6 per cent oxygen mixture for about 18 minutes) acted to decrease both sensory and motor conduction speed and the spinal and end-plate delays (Fib. 2). These opposed effects were not completely compensatory, so that overall reflex time tended to decrease. Renshaw suppression (Fig. 3) was considerably diminished, but returned to normal values when usual oxygen conditions were restored.

Hypoglycemia (a blood-glucose reduction of up to 0.07 per cent produced by insulin injection) provoked increases in all four of the reflex components studied (Fig. 4). These changes were almost completely compensatory with respect to total reflex time. When blood-sugar level was more severely lowered, motor conduction speed and end-plate delay increased still further, but sensory conduction speed and central delay showed an appreciable decrease. After-effects were observable for as long as 24 hours. Renshaw inhibition is considerably diminished by hypoglycemia.

Muscular fatigue (produced by one to one and a half hours of exercise on a cycloergometer) increased sensory and decreased motor conduction speed, the two types of change being approximately equal in magnitude. End-plate and spinal delay were usually decreased. Renshaw suppression was not studied under this condition.

Caffeine (250 mg given intravenously) had no effect on sensory or motor conduction speed, caused a slight increase in end-plate delay, and markedly reduced spinal delay, this effect being found to last for over four hours. An increase in the amplitude of the reflex response was also observed. Renshaw suppression, although severely reduced initially, was found after 80 minutes to have increased above normal, a level at which it remained for several hours.

Barbiturates (Diogenal given in doses sufficient to synchronize the EEG at the frequency of the alpha rhythm) produced no change in sencory or motor conduction speed, caused a slight increase in end-plate delay, and markedly increased spinal delay and Renshaw suppression. Response amplitude was increased and remained highly constant, a phenomenon observable in spinalized animals.

Alcohol (125 cc of a 95 per cent diluted solution administered in the briefest time possible) had virtually no effect on sensory conduction speed or end-plate delay. A small increase in motor conduction speed and a large increase in spinal delay were observed; these changes reached their maximum within the first half-hour and varied markedly in duration from subject to subject. Renshaw suppression was not studied.

Lack of sleep (loss of one night's rest) produced no change in sensory and motor conduction speed or in end-plate delay. Spinal delay was markedly increased and Renshaw suppression greatly reduced. (See Fig. 5). A full night's sleep was required to restore spinal delay to normal values, a brief period of sleep being insufficient.

The stress-inducing conditions studied may be grouped into two categories with respect to their effects on the dependent variables: (a) conditions which af-

fected both peripheral and central components of the reflex response, and (b) conditions which produced central changes only. Hypoxia, hypoglycemia, alcohol, and muscular fatigue fall into the first category; caffeine, barbiturates, and sleep loss fall into the second.

Ther peripheral changes (in neural conduction speeds and end-plate delay) could be due to direct effects on the nerve fibers themselves, or to indirect effects emanating from alterations in the functioning of the spinal cells. Direct evidence for the former hypothesis comes both from the present studies and from other research concerning the effects of hypoxia, hypoglycemia, and insulin on neural transmission.

The central effects observed under all stress-inducing conditions may be attributable to purely local effects, to supra-spinal influences, or to both. Insulin, hypoglycemia, and sleep loss are known to affect sub-cortical structures capable of influencing the activity of the spinal neurons, hypoglycemia having been demonstrated, in addition, to act directly at the spinal level.

With respect to the mechanisms mediating local effects in the cord, it is hypothesized that reduced speed of transmission at individual synapses and/or synchronization of synaptic activity in general could have contributed to the observed changes in spinal delay observed under insulin, caffeine, alcohol, and barbituric narcosis.

In conclusion, the need for data on the reflex changes produced by hormonal and psychological factors is emphasized.

Résumé

L'étude de la tension est malaisée parce que ce phénomène est d'une définition fuyante et les symptômes peuvent en être évoqués par des causes variées et se manifester selon des mécanismes divers. Certaines caractéristiques sont toutefois communes à tous les états de tension: (a) altération objective du fonctionnement et (b) fatigue subjective. On peut donc supposer l'intervention des certaines modifications du système nerveux central pendant la tension. Les présentes études se sont proposées de déceler des modifications de cet ordre susceptibles de se présenter dans diverses conditions de tension.

L'angle sous lequel cette étude a été abordée est celui des réflexes spinaux monosynaptiques—en raison de la facilité de mensuration chez l'homme de ces phénomènes fondamentaux, et de la traduction qu'ils semblent donner de l'état de l'ensemble du système nerveux central.

La procédure expérimentale adoptée repose sur l'évocation des réflexes spinaux dans la région sciatique par excitation directe, centralement et périphériquement en des points dont les interdistances sont déterminées de façon précise, d'un nerf spinal mixte (Fig. 1). Le temps de latence total ainsi que les intervalles entre excitation et réaction dans le muscle de la jambe, ont été mesurés et introduits dans

des équations convenables pour donner les indices des composantes suivantes du réflexe: (a) vitesse de propagation dans les fibres sensorielles, (b) vitesse de propagation dans les fibres motrices, (c) retard à la plaque motrice, et (d) retard spinal.

Le nerf a été excité par un stimulus électrique de 75 microsecondes à haute tension au moyen d'un appareil, spécialement construit, permettant un contrôle précis de l'intensité et de la durée du stimulus. Les erreurs dans les mesures relatives à la plaque motrice ont été écartées par le recours à une méthode normalisée de localisation des électrodes sur le muscle.

Un autre phénomène d'étude a été l'effet Renshaw, c'est-à-dire la courte inhibition d'un neurone moteur résultant de sa stimulation, immédiatement après la décharge, par une cellule internonciale de Renshaw. Les effets de la tension sur cette inhibition ont été analysés par comparaison de l'amplitude relative du premier et du second réflexe provoqué par deux stimuli appliqués au neurone moteur à 75 millisecondes d'intervalle (le temps exigé par le neurone pour récupérer, après la première excitation, environ la moitié de son excitabilité normale).

Les expériences comportaient les conditions de tension suivantes: hypoxie, hypoglycémie, fatigue musculaire, drogues (caféine, barbituriques, alcool), et manque de sommeil.

Les résultats expérimentaux peuvent se résumer comme suit:

Hypoxie (provoquée par respiration, pendant environ 18 minutes, d'air contenant y pour cent d'oxygène) a ralenti la propagation tant sensorielle que motrice, et a réduit les retards spinaux et à la plaque motrice (Fig. 2). Ces effets contraires ne sont pas tout à fait compensatoires, de sorte que le temps total de latence a tendu à decroître. L'effet Renshaw (Fig. 3) a été considérablement réduit mais a repris sa valeur normale avec la restauration des conditions ordinaires d'oxygène.

Hypoglycemie (réduction du taux de glucose sanguin jusqu'à 0,07 pour cent par injection d'insuline) a fait accroître les valeurs temporelles des quatre composantes du réflexe étudiées (Fig. 4). Ces modifications se sont compensées presque intégralement en ce qui concerne le temps total de latence. Après un abaissement plus sévère du taux du sucre dans le sang, la vitesse de propagation motrice et le retard à la plaque motrice ont augmenté encore plus, tandis que la vitesse de propagation sensorielle et le retard spinal ont subi un décroissement sensible. La persistance des effets a pu être observée pendant une période pouvant aller jusqu'à 24 heures. L'hypoglycémie réduit considérablement l'effet Renshaw.

Fatigue musculaire (engendrée par une heure à une heure et demie d'exercice au cyclo-ergomètre) a accru la vitesse de propagation sensorielle et a décru la vitesse de propagation motrice, ces deux altérations ayant approximativement la même grandeur. Les retards spinal et à la plaque motrice ont été généralement réduits. L'effet Renshaw n'a pas été étudié dans le cas de cette condition.

Caféine (injection intraveineuse de 250 mg) a été sans effet sur la vitesse de propagation sensorielle ou motrice, a fait accroître légèrement le retard à la plaque motrice, et a réduit nettement le retard spinal, ce dernier effet pouvant perdurer pendant plus de quatre heures. Un accroissement de l'amplitude de la réponse réflexe a également été observé. L'effet Renshaw, considérablement réduit au début, a paru, après 80 minutes, avoir dépassé au contraire sa valeur normale, atteignant un niveau où il se maintient plusieurs heures.

Barbiturique (administration de diogénal en doses suffisantes pour synchroniser l'EEG à la fréquence du rythme alpha) a été sans effet sur la vitesse de propagation sensorielle ou motrice, a augmenté légèrement le retard à la plaque motrice et a fait accroître nettement les valeurs du retard spinal et de l'effet Renshaw. L'amplitude de la réponse a été accrue et est restée très constante, phénomène constatable chez les animaux spinalisés.

Alcool (administration de 125 cc de solution diluée à 95 pour cent dans le délai le plus bref possible) a été virtuellement sans effet sur la vitesse de propagation sensorielle ou sur le retard à la plaque motrice. Un léger accroissement de la vitesse de propagation motrice et un important accroissement du retard spinal ont pu être observés; ces modifications ont attient leur maximum au cours de la première demi-heure, variant sensiblement en durée selon le sujet. L'effet Renshaw n'a pas été étudié.

Manque de sommeil (une nuit blanche) n'a modifié aucunement la vitesse de propagation sensorielle ni le retard à la plaque motrice. Le retard spinal a été nettement accru, et l'effet Renshaw sensiblement réduit (voir Fig. 5). Une bonne nuit de sommeil a été nécessaire au rétablissement de la valeur normale du retard spinal, un somme trop bref étant insuffisant.

Les conditions responsables de la tension étudiées peuvent être groupées en deux catégories quant à leurs effets sur les variables dépendantes: (a) les conditions affectant les composantes tant périphériques que centrales du réflexe, et (b) les conditions engendrant uniquement des modifications centrales. Hypoxie, hypoglycémie, alcool et fatigue musculaire se rangent dans la première catégorie; caféine, barbituriques et manque de sommeil dans la seconde.

Les modifications périphériques (dans la vitesse de propagation nerveuse et le retard à la plaque motrice) pourraient être dues soit à une action directe sur les fibres nerveuses mêmes, soit à une action indirecte engendrée par des altérations dans le fonctionnement des cellules spinales. A l'appui de la première hypothèse, les présentes études, aussi bien que d'autres recherches concernant les effets de l'hypoxie, de l'hypoglycémie et de l'insuline sur la transmission neurale, apportent un témoignage direct.

Les effets centraux observés dans toutes les conditions responsables de tension peuvent être imputables à des effets purement locaux, à des influences supraspinales, voire aux deux. L'insuline, l'hypoglycémie et le manque de sommeil sont connus pour affecter certaines structures subcorticales capables d'influer sur l'activité des neurones spinaux, la preuve ayant été faite par ailleurs que l'hypoglycémie agit en outre directement au niveau spinal.

En ce qui concerne les mécanismes intervenant dans les effets locaux dans l'épine dorsale, on suppose que la réduction de la vitesse de transmission aux divers synapses et/ou la synchronisation de l'activité synaptique en général ont pu contribuer à provoquer les modifications observées dans le retard spinal sous l'influence de l'insuline, de la caféine, de l'alcool et de la narcose par barbiturique.

En conclusion, la nécessité de données sur les changements de réflexe produits par l'action de facteurs hormonaux et psychologiques est souligné.

Commentary by Discussant

Chairman:

The discussant of this paper was to have been Dr. A. Fessard, of France, but, unfortunately, Dr. Fessard was unable to be with us today, the nearby Neurological Congress making stronger claims on him. Since there seem to be no participants who wish to offer comments, I shall ask Dr. Margaria if he would like to make any further remarks.

Reply

Dr. R. Margaria:

I should merely like to add that the experimental method we employ is a standardized procedure yielding highly reproducible results. It also yields measures of high accuracy, and is so painless that it can be used even with children.

It might be valuable to draw attention to some of the difficulties we have had in interpreting certain results. For example, we found that caffeine decreases spinal delay but increases muscular end-plate delay. This finding is rather difficult to rationalize, since neural transmission at the end-plate appears to be mediated by a synaptic process of the same type as that involved in transmission within the cord. Hence, it would be expected that a drug affecting one of these processes would affect the other in the same way and to approximately the same degree. Since the observed effects were, in fact, opposite in direction, it must be assumed that the decrease in spinal delay is apparently due not to a decrease in synaptic time, but to that known to be produced by caffeine in the spinal centers. This example illustrates that despite their apparent simplicity, the data require careful analysis of underlying mechanisms.

ENGINEERING FOR THE HUMAN

Julian M. Christensen
Psychology Branch, Aero Medical Laboratory
Wright Air Development Center
U. S. A.

Although the field of research covered in this paper is older in England than in the United States, it appears to be in the beginning stages of development on the continent. For this reason, it was decided to forego describing here the results of specific investigations and to present instead an overview of human factors research as it has developed and is currently being carried out in America. It is hoped that subjecting such research to the light of fresh perspectives will lead to a profitable exchange of views and information.

The term "human engineer" was omitted intentionally from the title of this paper. The term is not only somewhat unfair to human factors specialists and accredited engineers, it carries also the unfortunate implication that the human is being "engineered." Quite the opposite is true. Practically everything is "engineered" in this endeavor except the human.

However, the term has become popular, and those who work in this field are probably going to be called human engineers irrespective of their personal preferences. Terms such as engineering psychology, engineering physiology, engineering anthropology, etc., would be more accurate and more descriptive. A suitable generic term might be "human factors specialist." These specialists attempt to adapt tools, equipment, and the environment to men and to teams of men, taking cognizance of their individual and group characteristics, both inherent and acquired. The objective is to create a maximally effective system.

In addition to defining the area of work under consideration, this paper speculates concerning its origin and sketches briefly its recent history. The integrating effect this effort is having on related disciplines is noted. Certain deficiencies in the early work are discussed, and what is being done and planned in the way of correcting such defects is described. No specific design recommendations are made, such data being more completely and adequately covered in handbooks on the subject which are either available or in preparation. Because of the background of the author, somewhat more attention is devoted to engineering psychology than to other human factors areas, but this should not be taken as an indication of its relative importance.

A Further Definition of "Human Engineering"

The following is a gross description of the duties of human factors specialists. They must determine the nature and amount of information that a man or a team of men will require to make appropriate decisions. These decisions must be made explicit. They must ascertain how the required information should be dis-

594869 O - 61 - 5

played. Finally, they must determine the number and characteristics of the controls an operator should have in order to carry out decisions. In this work, their responsibility extends not only to the operational portion of a system, but also to the support portion and to the training equipment needed to provide the skills required for operation and support of the system.

In addition to these more obvious duties, human factors specialists must assist in the determination of those system functions which will be allocated to machines, those which will be allocated to men, and those which will be allocated to combinations of men and machines. These decisions must be related to the time required to accomplish the mission of the system, and to the time period during which the system is expected to be operational. A man-machine functions analysis for a system of ten years ago might appear ridiculous today.

The goal of human factors specialists engaged in systems design and development is completely compatible with that of the representatives of other scientific and engineering fields. Simply stated, this goal is to maximize the effectiveness of the system. However, human factors researchers have a unique and somewhat awesome responsibility, because in man they are faced with a system component which surpasses all others in terms of both potential sources of contribution and latent sources of error.

Reference has been made to the interdisciplinary nature of systems design and development problems. An interdisciplinary approach is not simply desirable—it is indispensable. The cold, hard problems of systems design do not come in neat, intradisciplinary packages. The design of a pressure suit, for example, is related to the design of controls and the layout of work places. These, in turn, are related to the design of the vehicle. No single group of engineering physiologists, engineering psychologists, or design engineers can contribute successfully to such a development without close coordination with several other groups.

Human engineering is perhaps best described as an "applied scheice." It is "problem-oriented," not "discipline-oriented." Perhaps it is presumptuous even to claim it is a science at all. However this may be, the practical value of the endeavor regardless of what it is termed or of how it is defined, is no longer in dispute. Moreover, there are indications that the social and biological sciences may ultimately benefit from their exposure in this arena to the thinking and procedures of other disciplines.

History and Early Developments

In a certain sense, human engineering is as old as man. The Einstein of the Cro-Magnon era was the being who decided not only to use objects, but to modify them. Over the succeeding centuries man made slow but steady improvements in most of his basic tools. These improvements, however, were based on intuition, or on trial and error, and not on a program dedicated to the systematic solution of design problems.

It is customary to refer to the occurrence of an unusually large number of changes in a short period of time as a "revolution." The era of the "Industrial Revolution is germane to this discussion. It was during this period that man learned to substitute machines for muscles.

In the laboratory, it is now called the "First Industrial Revolution," however, in the belief that the threshold of a Second Industrial Revolution has already been

crossed. This period will witness the wholesale substitution of machines for many functions that traditionally were considered "mental" in nature. By this is meant that the more ordinary of such functions will be performed by machines; the more challenging operations involving programming, detection and correction of malfunctions, judgment, etc., will be man's in even greater profusion and complexity.

Interestingly, around the turn of the century the engineer Frederick Taylor was on the verge of establishing human engineering, but he never quite succeeded. For example, he designed better shovels and other tools, and improved work methods. The Gilbreths followed with an emphasis on methods that still dominates industrial engineering; but, amelioration of equipment design—the ingredient in Taylor's heritage which could have thrust human engineering into the thinking of half a century ago—was more or less ignored.

Pursuing the logic of the job analysis approach, biological and social scientists in World Wars I and II placed their confidence almost entirely in the selection, placement, and training of personnel. Studies of equipment design were the exception, and studies of systems almost non-existent. It was apparent to many scientists in various countries, however, that personnel selection and training would never afford a complete answer, even for the airplanes of that era, and could not hope to deal successfully with the problems which would be posed by supersonic, super-altitude aircraft.[1] Similar problems were becoming apparent in the design in the design of land surface equipment, and sea surface and subsurface equipment.

Most of the early investigations in human engineering attempted to establish principles which could be applied directly to the solution of specific design problems. For example, considerable effort was devoted to investigation of the optimum size, shape, and direction of actuation of hand controls. Other studies involved traditional anthropometrics and functional sizing, determination of the limits of reach and manipulation, etc. Attention was given to the design and lighting of visual displays, and considerable work was done in the field of communications on such factors as redundancy, content, and carrier. Additional research concerned the design of training devices to yield a maximum of positive transfer, at minimum cost, from the training to the operational situation. These investigations were performed in accordance with the best scientific procedures and traditions, and form a basis for this science for which no one needs to apologize.

Limitations of the Early Work

Implicit in this early approach, i.e., attacking system design problems in a piecemeal manner, were several unresolved problems and questionable assumptions and techniques. First among these was the assumption that improvements in detail design ("knob and dial work," as it has been termed) would, when summed, yield a worthwhile overall gain. The possibilities of decrement (or, at least, of no significant improvement) due to unmeasured and unanticipated interactions were real.

---

[1] Agencies and names that deserve special mention here include, among others, Sir Frederick Bartlett, F. R. S., of the Medical Research Council, England; Dr. Paul M. Fitts, United States Air Force; and Dr. Franklin Taylor, United States Navy. Researchers in the United States feel a special sense of gratitude and indebtedness to Sir Frederic Bartlett who served as mentor and as worldwide leader of this movement.

Second, the experimenters frequently employed subjects whose representativeness with respect to the groups who would use the operational equipment was not known. Here, again, was a variable capable of producing significant changes in quantitative or qualitative relationships.

Third, important dimensions of behavior were often intentionally ignored, or at least not systematically controlled. Generally, this was due to a paucity of available information, and to difficulties in the definition, measurement, and control of these variables. Examples are stress, motivation, and fatigue.

Fourth and finally, was the familiar, omnipresent criterion problem. Of what do reliable and valid operational criteria of component and system effectiveness consist? Are there valid criteria sufficiently sensitive and isolable to permit assessment of the effects of human engineering changes?

Meanwhile, those engaged in actual design work were encountering further problems. Among these was the "too little, too late" complaint. Retrofitting is a time-consuming procedure in systems work, and applied researchers often found themselves addressing a problem so late in the development cycle that the worth of their suggestions was mitigated by considerations of time and practicability.

In addition, there was understandable concern regarding the extent to which one could extrapolate much of the experimental data available in the biological and social sciences. Surprisingly little information exists in the literature on this important problem of generality.

Present and Future Trends: Research

The criterion problem probably never will be completely resolved, but there are several promising approaches. Fitts suggests the following five criteria of effectiveness for human factors contributions to equipment and systems design: manpower, training time, efficiency, safety, and user acceptance. Somewhat related to these, but deserving specific mention, are the factors of dollar cost and time (delay).

Also involved in the criterion problem is an additional and more complex issue, namely, what are the relative weights or values to be assigned to the various measures of effectiveness? For example, how many units of performance (and please do not ask to have "units" defined here) can be sacrificed to gain $\underline{X}$ units of safety? How much additional time can be devoted to training in order to achieve greater efficiency? Such "trade-off" problems are characteristically multidimensional. Solving them, though not easy, is both necessary and possible. After all, design engineers implicitly assign values to different effectiveness criteria whenever they plan a system. In the final analysis, any system manufactured today is the result of a large number of compromises—compromists often made, however, on the basis of subjective weighting. It is important that such hitherto implicit weights and values be made explicit and submitted to objective study.

These and other considerations have stimulated investigations in "systems research." One of the frontiers of engineering psychology, this field deals with the fundamentsl principles governing the behavior of systems and with evaluation of system performance in terms of these principles. Systems research requires the simultaneous consideration of as many relevant variables as necessary in order to

In addition to the criterion problem, systems researchers are also confronted with what has been termed the "partitioning problem," i.e., the problem of identifying those parts of a system which can be extracted or abstracted for consideration without vitiating overall performance when the system is reassembled.

Then, too, there is the "description problem." What constitutes an adequate description of a system? Do the usual descriptions in terms of communication links, message content, subsystem input-output functions, etc., optimally throw into relief those operations of a system which are central to the interests of the human specialist?

These problems eventually will be solved by study and experimentation. Various criteria can be used, and their suitability and degree of overlap assessed. The partitioning problem can be attacked by treating it as an independent variable. The adequacy of various systems of description will become manifest through use. In these and other problem areas, logical analysis, experimentation, the construction and application of mathematical models, and the use of high speed computers and complex simulations will play important roles.

There are several groups which already can point to positive accomplishments in systems research. The Systems Development Corporation (formerly a division of the RAND Corporation) and the Aviation Psychology Laboratory of Ohio State University are examples of such groups. (It is with considerable pride that the Aero Medical Laboratory claims as alumni the chiefs of both these organizations—Dr. William C. Biel and Dr. Paul M. Fitts.) In addition to this effort, in-service people at the Laboratory are attacking the problem with vigor, and are gaining insights which will be highly useful in establishing experimental programs and solving operational problems.

Systems research has been considered here first both because of its inherent importance, and because eventually it may provide an "ultimate criterion" for the assessment of work in other human factors areas.

In examining these other areas, consider first the basic sciences on which human factors specialists so heavily depand—psychology, physiology, anthropology, sociology, and others. It is trite to say that more research is needed; the problem must be more specifically stated. Even more than additional basic data, the generality of the data already available needs to be known. How far and to what kinds of situations can existing information be reasonably extrapolated? The "basic" scientist might feel that this is a problem to which human factors specialists should address themselves, and perhaps he would be correct.

In the areas even more directly associated with design and development, is the field of "applied experimental" work. In the area of equipment controls, modest success has been experienced in writing transfer functions for the human motor-response system. Eventually, this endeavor may enable design engineers to specify mathematically the process characteristics of the human operator as a component of a system. Meanwhile, continual attention is being devoted to shape coding, size, effects of work-space restriction, and other dimensions of control design.

An immediate challenge in the visual display area involves the determination not only of "how," but also of "how much." Mackworth has called attention to this general problem in his excellent Sir Alfred Herbert Paper for 1955, "Work Design and Training for Future Industrial Skills." He points out that modern computers

not only can turn out appalling stacks of figures, they can also select and summarize. He feels that this attribute of computers has not been adequately exploited. Mackworth proposes discovering how little information must be recorded and processed in order to obtain a desired result, or, at least, how little should be presented finally to man for treatment. In this, the author heartily concurs and would add only that the decision regarding what information finally to present to man and what control he will exercise will vary widely under different conditions.

In addition, more information is needed on the nature of "sensory overload." Many have assumed that the factors limiting man's ability to receive and process information are defined by visual scanning rate, number of absolute or relative discriminations which can be made, etc. Suggested solutions, therefore, usually involve "relieving" the overburdened sense modality by presenting information to one or more of the other modalities. But, will this approach lead to the best solution? Perhaps the limits are defined in the central nervous system. If so, Mackworth's plea is even more poignant.

Those assisting in the design of training equipment for weapon systems also face new and fascinating problems. For example, how does one train for expertness in monitoring and decision-making—functions the human operator will be asked to perform more and more frequently in the future? How should one transmit the knowledge of results that is so essential to efficient learning and continued motivation? And how should a man be trained so that he can maintain efficiency throughout long periods of monitoring?

Very little mention has been made of human factors in the area of communications. This is left in the capable hands of Dr. Karl Kryter, who will deal with communications problems in his paper.

Finally, there is the problem of space flight. One day, man will take his first step into space. At that time, sound advice must be available on the behavioral effects of increased and decreased acceleration and attendant conditions of weightlessness. Equally important, estimates must be made regarding the effects of prolonged detachment from the physiological, psychological, and social environments of the earth.

Present and Future Trends: Applications

In the application of human engineering data, there are problems other than those relating to basic information. Of initial importance is the problem of timeliness, to which reference was made previously. Human factors contributions must be consonant with the development schedule of a system. The Aero Medical Laboratory has had a contractor develop a procedural guide for design engineers and human factors specialists. This document traces in detail the development schedule of a weapon system. It enumerates those stages at which specific human factors decisions must be made, and specifies the information that must be available to make such decisions in the design of Air Force systems.

Related to this question is the responsibility human factors specialists must bear for their decisions. Manifestly, it should be similar to that which any other expert has to the chief design engineer. The days of offering expert advice from a cloister are, fortunately, almost at an end.

The human factors specialist must maintain a broad perspective with respect to the ultimate criterion—system effectiveness. He must recognize and constantly remind himself that this refers to all aspects of the system—ground support and personnel utilization, as well as efficient operational design.

In conclusion, it can be mentioned that there is as yet no adequate handbook of human factors data for use directly by design engineers. However, the three services of the United States will soon publish a comprehensive guide to available information relating to controls, control dynamics, displays, anthropometry, workplace layout, etc.

Conclusions

Human factors considerations must receive adequate and timely attention if modern systems are to be maximally effective.

Research is needed especially on methods of systems design and structure, training equipment design, and the effects of space travel on behavior. Investigations are required that will indicate where, when, and how far one can extrapolate from existing basic data. Information is needed that would result in the generation of "trade-off curves" with respect to criteria of effectiveness.

The human factors applications specialist must take a continuing, active interest throughout the planning, design, development, testing, and operational use of systems. He must assume his fair share of responsibility for the success or failure of this very complex undertaking—the creation of modern man-machine systems.

Appendix: Some Suggested Human Factors Research Areas[2]

Controls

1. Research on the human factors design criteria of remote control devices. Information is needed concerning the behavioral effects of such factors as the nature and fidelity of feedback, and the constraints and limitations on modes of operation.

2. Research on simultaneous discrete and continuous control operation. Information is needed on the human operator's performance when he is engaged simultaneously in a continuous operation and in auxiliary discrete tasks.

3. Research on the derivation of nonlinear mathematical transfer functions for the human operator. Previous work in this area has been concentrated on the derivation of linear descriptive functions, and has lumped the remaining uncorrelated output. Consideration of the human operator as a nonlinear system is needed.

Space Psychology

1. Effects of sub-gravity conditions on human behavior. What are the effects of the weightless state on performance?

[2]Suggested by staff members of the Psychology Branch, Aero Medical Laboratory.

2. Effects of increased gravitational forces on performance. Investigations are needed conderning man's ability to carry out various actions under increased gravity.

3. Display of information in space vehicles. Some of the conventionally displayed parameters, e.g., altitude, will be meaningless in space flight. However, orientation with respect to some meaningful point or points of reference will no doubt be necessary. Exactly what will be required is a point for serious study.

4. Duties of space crews. In addition to research on informational requirements, careful study of the nature of the duties of space crews is needed.

5. Space suit design. Studies need to be performed regarding the nature and amount of the restrictions imposed on man by protective suits usable both within and without the vehicle.

6. Long periods of confinement. Living quarters in a space vehicle necessarily will be small and confining. Information is needed regarding the behavior of small groups during long periods of close confinement.

7. Diurnal cycles and behavior. In space, time zones are measured in astronomical rather than in earthly time. More information is needed concerning the effects on behavior of serious disruption of the diurnal cycle.

8. Training equipment. The training equipment requirements to prepare a crew for space flight need to be thoroughly analyzed.

Stress and Fatigue

1. The development of electrophysiological measures of performance. Developments in this area would add precision and reliability to studies involving the expenditure of energy, and would provide sensitive indices of the effects of stress on performance.

Systems Research

1. There is an urgent need for a comprehensive set of terms defining the behavior of the human operator in a systems setting. What, for example, is really meant by "decision process"? How can we develop a unifying basis of classification for the various kinds of behavior? Answers to such questions are needed to provide a firm semantic foundation for other investigations.

2. There is a need for the development of new concepts of behavior. These concepts should be meaningful for the systems investigator, and should admit of operational definition for application to systems experimentation. Furthermore, they should provide a link with classical psychological knowledge, being unique but not foreign to the tradition of psychology.

3. The development of new theories for systems research, including fresh approaches to the problems of description, analysis (including "partitioning"), and synthesis.

4. Research in mathematical model-building, with special emphasis on the solution of complex problems involving higher-order interactions.

5. The development of reliable and valid criteria for assessing the value of new controls, displays, and cockpit configurations. Consideration should be given to the various types of aircraft and missions involved. Much of the initial work might be done in ground simulators.

Training Equipment

1. Create a widely applicable taxonomy of tasks and behaviors. The purpose would be to (a) facilitate the organization of existing information about factors which influence the learning, transfer, and retention of specific classes of behaviors, and (b) aid the search for basic principles which will allow reliable generalization from existing experimental data to a wide population of laboratory and real life tasks.

2. Investigate the influence of learning one task upon the learning of another afver varying periods of time. Determine how "time between tasks" affects some of the factors, e.g., degree of similarity which influence transfer of training, giving due consideration to the fact that various aspects of a task may be differentially affected.

3. Carry out research aimed at determining the extent to which existing knowledge about learning, transfer, and retention can be generalized to various conditions, tasks, and subject populations.

Visual Displays

1. Basic research on task variables involved in the decision-making process. Attention might be given to such independent variables as the number of information dimensions, the amount of information per dimension, methods of sensory presentation of the information, and the number and kind of response alternatives available.

2. Visual perception of complex forms. An attempt should be made to isolate the variables affecting a subject's ability to discriminate forms. Particular emphasis should be directed toward systematizing and quantifying those aspects of a form which make it discriminable from other forms.

3. The "attention-getting" value of various stimuli. Determine, for various conditions of masking and primary task load, the characteristics (frequency, intensity, etc.) of those stimuli which are effective in redirecting attention from the primary task.

4. Use of the tactual sense as a channel for information input to the human operator. It is important to have information concerning the relative merits of different sense modalities for receiving various types of information.

Summary

The paper is devoted to: (a) defining the nature and purpose of research on human factors in relation to systems development and design, (b) tracing the history of such work, (c) describing deficiencies in early approaches and how they are being corrected, and (d) drawing attention to current problems and needed research. It is suggested that workers in this field be referred to as "human factors specialists" rather than as "human engineers," the latter term implies that the focus of engineering attention is on the human operator rather than on the equipment he employs.

The purpose of human factors research is to adapt tools, equipment, and the environment to men and teams of men in such a way as to create a maximally effective system. This work includes: (a) determining the amount and type of information needed for appropriate decision-making by the human operator: (b) developing effective methods of information display and efficient equipment controls; and (c) deciding which systems functions should be allocated to men, to machines, and to man-machine combinations. The researcher's responsibility extends to training equipment and to systems support, as well as to operational problems. Human factors research is an applied science which is, by nature, interdisciplinary.

Historically, the origins of the discipline can be traced to man's first attempt to improve his tools. Around the turn of the century, Frederick Taylor gave systematic attention to this problem, but it was his research on the improvement of work methods which received subsequent emphasis. During World Wars I and II the "personnel approach" (selection, placement, and training) dominated the field.

In contrast to the glovel approach of modern "systems research," early studies in the human factors area concentrated on more specific and isolated design problems. Deficiencies in earlier studies included: (a) assuming that improvements in detail design would increase overall system effectiveness, (b) utilizing subject samples of unknown representativeness with respect to populations of operational personnel, and (c) ignoring important but difficult-to-measure aspects of behavior, e.g., stress, motivation, fatigue. Among problems still current are those of (a) defining satisfactory criteria of system and component effectiveness, (b) insuring timeliness of design suggestions, and (c) determining the extent to which existing experimental data can be extended to other situations.

As regards present and future trends, it is noted that current emphasis is on "systems research," i.e., concentration on the principles governing the overall functioning of man-machine combinations. Possible criteria of system effectiveness include: manpower, training time, efficiency, safety, user acceptance, monetary cost, and time needed to make suggested design changes. The formulation of multiple criteria, however, raises the problem of what relative value to assign to each, and of how to effect optimal compromises between them.

Other current research problems include: (a) the "partitioning problem" (how to select parts of a system to be isolated for separate study without vitiating overall system effectiveness), and (b) the "description problem" (how to choose the most useful method of describing systems for purposes of research).

Human factors areas other than systems research include: (a) basic research in the behavioral sciences (where there is a need both for more empirical data and

for knowledge of the degree to which existing findings can be generalized); (b) the design of equipment controls (in which progress has been made in describing mathematically the responses of the human operator as a component of a system); (c) the design of visual and other displays (which involves deciding how much information should be provided to the human operator, as well as how it should be presented); (d) the design of training equipment for weapons systems (including the problems of how best to transmit knowledge of results, and of how to train for efficient decision-making and monitoring); (e) communications research (to be treated in a subsequent Symposium paper); and (f) research on the behavioral problems posed by space flight (effects of increased and decreased gravity, of isolation from familiar psychological, physiological, and social environments, etc.).

In discussing the application of research findings, the author re-emphasizes the importance of making human factors design suggestions at appropriate moments in the systems development cycle. It is stressed that the human factors specialist has a responsibility to the design engineer equal to that of any other consultant, and that his contribution to maximization of system effectiveness extends to ground support and personnel utilization as well as to design problems.

A section listing needed studies in various human factors research areas is appended.

Résumé

Dans cette communication, on se propose de (a) définir la nature et l'objet des recherches portant sur les facteurs humains dans le cadre de la conception et de la mise au point de systèmes; (b) retracer l'historique de ces travaux; (c) décrire certains défauts des premières études et les correctifs qui y sont apportés; et (d) attirer l'attention sur les problémes courants et les travaux de recherche nécessaires. Il est suggéré que les chercheurs se consacrant à ce domaine devraient être appelés "spécialistes des facteurs humains" ("human factors specialists") plutôt que "human engineers" (c'est-à-dire "ingénieurs de l'humain"); cette dernière expression implique que l'étude porte sur l'opérateur humain plutôt que sur l'équipement qu'il utilise.

Les recherches sur les facteurs humains ont pour objectd'adapter les outils, le matériel, et le milieu aux hommes et aux équipes de telle sorte qu'il en résulte un système efficace au maximum. Ce travail comporte: (a) une détermination du type et de la quantité d'information requise pour permettre à l'opérateur humain de prendre les décisions appropriées; (b) la mise au point de méthodes effectives de présentation d'information et de contrôles efficaces du matériel; et (c) le choix des fonctions à conférer aux hommes, aux machines et aux complexes homme-machine. La responsabilité impartie au chercheur embrasse l'équipement de formation et l'appui des systèmes aussi bien que les problèmes opérationnels. La recherche sur les facteurs humains est une science appliquée essentiellement interdisciplinaire.

Si l'on retrace l'historique de cette discipline, on peut en faire remonter les origines aux premiers efforts de l'homme pour améliorer ses outils. Au tournant

du siècle, Frederick Taylor a étudié ce problème systématiquement, mais ce sont ses recherches sur le perfectionnement des méthodes de travail qui, dans la suite, retinrent le plus l'attention. Au cours des deux guerres mondiales, le champ était dominé par les études portant sur le personnel (sélection, placement, et apprentissage).

Contrairement au point de vue global moderne de la "recherche de systèmes," les premières études sur les facteurs humains se concentraient sur des problèmes plutôt spécifiques. Leurs défauts peuvent se caractériser comme suit: (a) admettre a priori qu'une conception perfectionnée du détail devait accroître l'efficacité générale du système; (b) opérer à partir de sujets dont la valeur représentative des populations de personnel opérationnel était indéterminée; et (c) ignorer certains aspects importants mais difficilement mensurables du comportement (par exemple tension, motivation, fatigue). Parmi les problèmes encore actuels sont ceux qui consistent à (a) definir des critères satisfaisants de l'efficacité du système et de ses éléments; (b) assurer l'opportunité des suggestions conceptuelles; et (c) determiner la mesure dans laquelle les données expérimentales disponibles peuvent être étendues à d'autres situations.

En ce qui concerne les tendances tant courantes que futures, il est noté qu'actuellement l'accent est mis sur la "recherche de systèmes," c'est-à-dire sur les principes régissant le fonctionnement global des complexes homme-machine. Comme critères possibles de l'efficacité des systèmes l'on peut citer: l'énergie humaine ("manpower"), le temps d'apprentissage, le rendement, la sécurité, l'acceptation par l'utilisateur, le coût financier et le temps nécessaire à faire des changements conceptuels suggérés. La formulation de critères multiples pose, toutefois, le problème de la valeur relative à attribuer à chacun d'eux, et de l'optimalité des compromis à établir entre eux.

D'autres problèmes courants en matière de recherche comprennent: (a) le "problème du compartimentage" (comment scinder un système en parties susceptibles d'études distinctes sans que l'efficacité du système global en soit affectée) et (b) le "problème de description" (comment choisir, aux fins de la recherche, la méthode la plus utile pour décrire les systèmes).

D'autres domaines de recherche sur les facteurs humains comprennent: (a) la recherche fondamentale dans les sciences du comportement (où sont nécessaires plus de données empiriques et une connaissance plus approfondie de la mesure dans laquelle les découvertes acquises peuvent être généralisées); (b) le perfectionnement de contrôles du matériel (où des progrès ont été réalisés par la traduction mathématique des réactions de l'opérateur humain en tant qu'élément actif d'un système); (c) l'élaboration des méthodes de présentation d'information (y compris la détermination tant de la quantité que de la forme de l'information à fournir à l'opérateur humain); (d) l'élaboration des équipements de formation aux systèmes d'armements (y compris les problèmes de la communication optimale des résultats et de la préparation à des responsabilités de décision et de surveillance); (e) recherche sur les communications (qui fera l'objet d'une communication ultérieure dans le cadre de ce symposium); et (f) la recherche sur les problèmes de comportement posés par les vols spatiaux (effets de l'accroissement et du décroissement de la gravité, de la coupure du milieu psychologique, physiologique et social habituel, etc.).

Discutant l'application des résultats de la recherche, l'auteur souligne une fois de plus l'importance du choix du moment opportun, dans le cycle de développe-

ment des systèmes, pour présenter des suggestions conceptuelles en matière de facteurs humains. La responsabilité du spécialiste en facteurs humains envers l'ingénieur dessinateur est, souligne-t-on, égale à celle de n'importe quel autre consultant, et son apport à la maximalisation de l'efficacité du système va depuis la formation et l'utilisation du personnel jusqu'aux problèmes conceptuels.

Une annexe relève les diverses études nécessaires dans différents domaines de recherche en matière de facteurs humains.

Commentary by Discussant

Mr. Hugh C. W. Stockbridge (U. K.):

It gives me great pleasure to introduce the discussion of Mr. Christensen's original and provocative paper. I was particularly interested by his comments on systems research, which forms a part of our program at the Experimental Establishment. Some of the thoughts occurring to my colleagues and myself have been quite comparable to those expressed by Mr. Christensen. In systems research many of the individual components are often given elaborate laboratory tests—for example, the lives of valves are carefully computed—but the experimental method is not always adequately carried over to the control display. The guiding principle toward which we have tried to work is that all the components of a system should be experimentally tested by appropriate methods.

As regards criteria of effectiveness, I tend to take a somewhat narrow view and to think mainly in terms of time, at least with respect to systems which are self-correcting. If we imagine a number of men seated at desks passing messages to each other, and one of them makes a mistake, he will correct his error (or so we hope). But, this correction will take time. It is possible that even such criteria as safety might ultimately be measurable in terms of time; training, of course, certainly can be. Although perhaps rather narrow, this view does have the advantage that time is an easy thing to measure.

Mr. Christensen has also mentioned the problem of interactions among variables. A recent study carried out at the Establishment provided an interesting example of this. It was found that young male subjects have lower body temperatures when measured individually than they do when measured in groups. This finding raises the interesting general question of whether other basic physiological variables reflect such changes, and whether the correlations and interactions among them could be used as objective measures of such things as group cohesiveness and morale. Questions of this kind also fall within the broad province of systems research.

In discussing "trade-offs" and compromises among criteria of effectiveness, Mr. Christensen appeared to conceptualize this problem, quite correctly I believe, as essentially one of attempting to maximize the various possible criteria. However, in taking such a very broad view of the problem, I believe one runs the risk of straying far outside the province or powers of the psychologist proper. Any at-

tempt to maximize all the possible criteria of effectiveness—to arrive, as it were, at the top surface of this n-dimensional space—necessarily leads to the question of whose job it will be to carry out such maximization. This question leads ultimately into the realms of politics and finance. I believe that in operations research the commonly accepted view is that one may advise, but should be extremely careful to go no further. An escapist outlook, perhaps, but one which many of us, as scientists, are forced to take.

In conclusion, I should like to say that I whole-heartedly agree with the author's remarks about the many areas of research in engineering psychology which have not yet been touched, or in which a great deal more work could be done. For example, the interaction of the various forms of environmental stress occurring in battle is a problem of great interest to us. There are many others. I was particularly interested by Mr. Christensen's statement of the numerous problems which will be posed to the psychologist by developments in the field of space flight.

General Discussion

Dr. Ward Edwards (U. S. A.):

I wish to speak about only one of the many issues raised in Mr. Christensen's extremely interesting and far-reaching paper, namely, the issue of quantitative relationships among different criteria—the issue, for example, of how much extra training is worthwhile in order to prevent one accident. Mr. Christensen has requested us not to ask him what units he had in mind when he talked about "one unit of training." However, I think the problem might appear a little simpler if we consider exactly this question.

It happens that for the last thirty or forty years the problem of units of value or desirability has been a central one in economic theory. Economists have invented a name for measures of value: they call them "utilities." They have done a great deal of work on the general problem of interacting values and have come up with some mathematical results which I think the systems researchers in psychology use less often than might be advantageous. One of these results is that if you consider only riskless cases, i.e., cases in which it is known for certain what the outcome of a chosen policy will be, then "measures" of value are not really needed at all. You need only what are called "ordinal scales," that is, statements as to whether one situation is preferable to another. Moreover, given such ordinal data, is is possible to combine them by means of a device called an "indifference curve" in such a way as to arrive at an optimal criterion for action—a criterion based on several complex, interacting variables.

Of course, if it is necessary to consider the more common situation in which the outcome of a given policy is not certainly known, then interval measures, i.e., measures defined up to a linear transformation, must be utilized. However, the fact that two points on any interval scale may be chosen arbitrarily means that even in risky situations nothing more than ordinal information about values may be needed. If only two points on the value scale are necessary, they may be taken to be the two

arbitrarily assigned points permitted by interval scales. If more than two points are necessary, additional points need be evaluated only with reference to the two arbitrary points. It is rare that many different points on any value dimension are required; a small number of categories, e.g., good and bad, is usually enough.

In brief, then, mathematical techniques available in the literature of economics deal precisely with the question of interacting dimensions of value, and with how such dimensions can be combined to permit identification of an optimal policy or course of action. Human engineers might find these techniques highly useful in dealing with criterion problems of the type described by Mr. Christensen.

Dr. Christine Kris (U. K.):

I should also like to comment briefly on this problem of multivariate relations. At the Massachusetts Institute of Technology, where I am presently working, a group of researchers has been concerned with the study of physiological and neurophysiological functioning from the standpoint of general systems theory. Among the techniques they employ are statistical methods appropriate to the analysis of multivariate relations, both across time and across variables. Some of these methods appear to be appropriate to problems of the type described by Mr. Christensen. It would, therefore, seem that systems researchers might profit from an examination of these methods as well as from study of the econometric techniques mentioned by Dr. Edwards.

Mr. Donald Wallis (U. K.):

I should like to put to Mr. Christensen a question about the economics of human engineering. In the United Kingdom, we have been working with engineering psychology problems for some years, but the area of systems research has been inadequately probed. The main difficulty is that we do not have enough trained personnel to analyze our major new systems to the extent that human engineering principles demand. I should welcome some indication from Mr. Christensen as to how he thinks our available supply of personnel might best be trained and deployed.

Dr. James C. Penton (U. K.):

I should like to put a question to Dr. Christensen. Owing to the very small number of psychologists available for this kind of work, it is a very great problem to us to know how the training of the people who are going to work in human engineering should be directed. There are, I suppose, two or three possibilities. You have three different disciplines impinging on this field: the psychological, the engineering, and the physiological. Ideally, one should have teams, presumably consisting of people from all disciplines working together, but this is not possible, at any rate in the United Kingdom, where the supply of people is quite inadequate. Are we then to send some of our psychologists to engineering a while so that they can learn sufficiently about that discipline to talk sense to their colleagues; or, are we to try to teach the engineers sufficient psychology and physiology so that they can, with psychologists and physiologists as consultants, solve their own problems? I would be very grateful if Dr. Christensen could give me any idea as to the lines along which that problem is being solved in the United States.

Dr. Alphonse Chapanis (U. S. A.):

At Johns Hopkins University, training in human engineering can proceed either through psychology or through industrial engineering. In fact, I have held a joint professorship covering both disciplines. Students are accepted from either field, and the graduate training includes subject matter of the following categories: operations research, psychology, time-and-motion study, fundamentals of electrical engineering, fundamentals of mechanical engineering, mathematics, statistics, environmental climatology, and applied anthropology.

Reply

Mr. Julien M. Christensen:

With respect to the comments of Mr. Stockbridge concerning the psychologist's responsibility for the maximization of system effectiveness, I should like to say that I personally see the contribution of the engineering psychologist as unique, and as constituting only a small part of the overall research effort. The systems research team in our own laboratory already includes a social anthropoligist, a mathematician, and two people with training in both engineering and psychology. So, it can be seen that in practice as well as in theory we take an interdisciplinary view of this work.

The above comments relate also to the question raised by Mr. Wallis and Dr. Penton. In the United States, we also lack people with sifficiently diversified training to handle the many different kinds of problems posed by systems analysis and development. I therefore believe that the personnel problem in this area will have to be solved partly through interdisciplinary teamwork.

Multidisciplinary training, however, can also make a contribution. Dr. Chapanis has described for you the comprehensive nature of the human engineering course content at Johns Hopkins University. Ohio State University has a comparable arrangement. There are still other universities which offer similar instruction. Thus, one can say that in the United States we are attacking this problem on two fronts. In the long run, both approaches may prove equally necessary to progress.

SESSION II

Chairman: Lt. Col. Charles Chandessais

Director of Human Sciences

Permanent Secretariat of the Scientific Action Committee of National Defense
(Secrétariat Permanent du Comité d'Action Scientifique de Défense Nationale)

Paris, France

594869 O - 61 - 6

A SPECIAL APPROACH TO ANALYSIS OF THE FUNCTIONS OF THE NON-COMMISSIONED OFFICER

Louis Delys
Center of Psychotechnical Studies and Research
Belgian Armed Forces
Brussels, Belgium

Several years ago, the Belgian Army, then in the process of rapid expansion, was faced with the problem of selecting a large number of non-commissioned officers. An investigation aimed at specifying the functions of the NCO was launched, therefore, at the Center of Psychotechnical Studies. The study involved conducting a survey among unit and training-school commanders.

In the survey, information was requested on the causes of failures among reserve NCO's in training and on the job. Analysis of the results revealed wide differences between the criteria used to define success in training school and the criteria used to define success on duty in the units. The school commanders were concerned primarily with the intellectual and physical aptitudes of the candidates, whereas the unit commanders laid stress solely on practical aspects, and, in some cases, even went so far as to criticize the schools for failing to provide NCO's with professional training and education in the handling of men.

It was concluded, however, that the differences in criteria were more apparent than real. The unit commanders come into contact with the NCO only after he has passed through training school, and have a quite natural tendency to minimize the importance of those attributes which are already possessed. The differences in viewpoint between the units and the schools can therefore be attributed to the fact that their respective aims are successive rather than simultaneous.

It was accordingly agreed that a new approach to the problem must be undertaken. Since the selection procedures already in force specified the indispensable minimum requirements for success in training school, it was decided to concentrate on an attempt to identify characteristics related to the success of reserve NCO's in carrying out their duties on the job.

Study of the literature on group discussions and the development of the critical incident technique suggested that a synthesis of these two methods might provide more complete and objective data than could be otherwise obtained. Various groups of unit commanders were invited, therefore, to participate in discussions devoted to an analysis of non-commissioned officer qualities.

Group Discussion Phase

In December and January, 1954, a series of eleven group discussions (two for the infantry, two for the artillery division, two for the service and transport corps, two for the engineer and signal units, two for the supply and administrative

services, and one for the armored units) was organized with the aim of gathering first-hand information concerning the job requirements of NCO's. Each of the discussions was attended by eight unit commanders from the various service branches specified. Three psychologists and the author acted as observers at all discussions which they helped to animate, but in which they did not otherwise participate.

The discussions were divided into two sections: A "structure" discussion, held in the morning, and an "unstructured" session, involving the same participants, held in the afternoon. In the structured discussion, use was made of the critical incident technique. The group members were asked to describe the behaviors manifested in various situations by: (a) particularly good and bad NCO's, and (b) individuals who could be expected to develop into either good or bad NCO's. It was requested that all remarks be based on actual cases drawn from personal experience. In the following unstructured discussion, the participants described, again in terms as concrete as possible, the chief overall qualities which they considered to be requored of a good reserve NCO. Also discussed were the estimated value of these observations and the conclusions of the previously conducted survey.

All instances of specific behaviors cited during the discussions were recorded at the time by the observers. The resulting records yielded a very extensive sample of items. To reduce the sample to manageable size, a classification of the items was carried out and a list of 50 typical behaviors was derived.

Evaluation of the Group-Discussion Data

The list of 50 behaviors thus obtained was sent to 14 school commanders, 17 training-center commanders, and 265 units commanders for grading into five classes with respect to judged importance. It was requested that the items be assigned to the classes in such a way that the categories ranging from most to least important would contain, respectively, 5, 10, 20, 10, and 5 of the total number of behaviors listed.

The data of that portion of the study were analyzed in two ways. In the first analysis, comparisons were made among the complete classifications received from all three of the officer groups contacted. In the second, attention was restricted to comparing the behaviors assigned to the first two classes by the unit commanders (and thus considered by them to be of "very great" or "great" importance), with the appraisals made by the school and training-center commanders.

The results of the first analysis, in which those behaviors receiving unanimous approval by all of the raters were picked out, revealed the existence of very great differences of opinion among the three types of commanders regarding the relative importance of various officer qualities. Each arm or service appeared to require its own specific pattern of physical, intellectual, and personality characteristics.

Results of the second analysis tended to confirm the findings of the first. Only eight of the behaviors classified as of great or very great importance by the unit commanders were found to be applicable to all the arms and services. Moreover, certain of these behaviors could be considered to represent components of the same basic trait, and could be linked to four general qualities of the successful NCO, i.e., ability to assimilate professional training, assertiveness of personality, ability to identify with the group, and devotion to duty. In general, the results of that phase of the study showed that the patterns of requirements for each arm and service are so different that a generalized picture of the good reserve NCO is difficult to draw.

Grouping of the Items into Clusters

To facilitate the study of characteristic differences in NCO requirements among the various arms and services, it was decided to attempt to classify the 50 behavior items into psychologically coherent clusters.

The items of the list of 50 behaviors were therefore analyzed in accordance with the values attributed to them by the raters. The items retained from this study consisted of: (a) those which had been classified in the first two categories ("very important" and "important") by the unit commanders, and had not been given a poor rating by the school and training school commanders, or by those unit commanders who had taken part in the group discussions, and (b) those which had been classified in the third category ("average" importance) by the unit commanders, and had been considered to be of at least average importance by the school and training center commanders, as well as by the group discussion participants. Items from the lowest two of the original five categories were discarded.

From the items retained, a set of 12 clusters was composed (see Appendix for list). The grouping of the items was done on a subjective basis, and some behaviors were placed in more than one cluster. In making up the clusters, it was necessary (although admittedly questionable) to assume not only that the items were of equal importance, but also that they were independent of one another.

Each of the resulting clusters was then assigned a weight, based on the number of items it contained, in such a way as to yield a total score across clusters of 100 points. The resulting set of 12 values was considered to represent an overall grading system applicable to NCO's in general, and thus to constitute a "basic reference pattern," indicating those qualities presumably required of NCO's by all the arms and services.

The basic reference pattern having been obtained, a further analysis of the items was carried out in order to identify the patterns of qualifies specific to each arm and service. In the analysis, all the behaviors included in the clusters were classified into three groups according to the degree of agreement as to the items' importance found among all of the raters who had taken part in the investigation. The first group contained items about which there had been general agreement, the second contained items on which the majority of the raters had agreed, and the third comprised behaviors whose importance was debatable. The items in those three groups were then assigned weights of 4, 3, and 2, respectively, and appropriate weighted sums were calculated for each of the previously derived clusters (again setting the total at 100). This was done for each arm and service studied. A table resulted which contained, for each branch, a set of 12 weighted values which could be regarded as representing the specific pattern of NCO qualities typically required by that branch. Table 1 presents an excerpt from this general table; it shows the basic reference pattern together with the specific patterns derived for the infantry and the administrative service.

Inspection of the various sets of values contained in the main table provided information on two general questions. First, comparison across the specific patterns served to identify those NCO characteristics which were actually considered important, in greater or lesser degree, by all the arms and services. (It was found that, with the single exception of Cluster II for the administrative service, the traits represented by Clusters I, II, III, IV, VI, IX and XII are required to varying extents by all the branches studied, the remaining qualities being considered important only

TABLE 1

EXCERPT FROM GENERAL TABLE OF ITEM-CLUSTER VALUES SHOWING BASIC REFERENCE PATTERN AND SPECIFIC PATTERNS FOR THE INFANTRY AND ADMINISTRATIVE SERVICE

Cluster	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Tot.
Reference pattern	20.0	12.5	13.75	7.5	7.5	11.25	5.0	3.75	6.25	3.75	2.5	6.25	100
Branch-specific patterns:													
Infantry	11.4	21.7	8.0	18.1	12.5	3.4	9.1	----	9.0	--	--	6.8	100
Administrative service	25.6	----	26.8	3.6	--	8.5	--	12.2	13.5	4.9	--	4.9	100

by certain branches.) Second, comparison of the specific patterns with the basic reference pattern provided estimates of the relative degree of importance assigned, within the various arms and services, to each of the traits represented by the clusters.

In summary, then, it can be said that the results of the analyses provided a means of estimating: (a) which qualities tend to be required of the NCO by all arms and services, (b) which traits tend to be more or less specific to certain arms, and (c) the comparative degree to which a given quality is valued in each of the various branches.

Conclusions

1. Owing to differences in the qualities required by the various arms and services, it is virtually impossible to formulate a completely generalized definition of the NCO, i.e., of a subordinate officer who would be equally successful in any branch of the service.

2. It is possible, however, to specify certain qualities required in common (although in varying degrees) by all service branches. These characteristics include: suitability for instruction and training, motivation for and devotion to duty, ascendancy or personal assertiveness, sociability, initiative, practical intelligence, and relevant knowledge and skills.

3. Certain officer requirements tend to be specific to certain branches of the service. In this regard, the following findings deserve mention. (a) A relatively high degree of general intelligence and education are required of the NCO by only a few arms and services, e.g., artillery and signals divisions. (b) Special training prior to the taking up of duties is considered of appreciable importance only by the supply and administrative branches. (c) Vitality and a good physique are required only in the combat services.

In general, the findings suggest that selection of non-commissioned officers should be conceived in differential terms, and approached from the standpoint of specific branch requirements.

Appendix: Clusters of Behavior Items

I. Motivation for duty: liking for and satisfaction with job; conscientiousness in the care and maintenance of equipment; use of methodical work methods; sense of authority; soldierly attitudes; esprit de corps; sense of social, civic, and professional responsibility.

II. Ascendancy: assertiveness; ability to compel recognition of personal authority, to dominate the group, to secure obedience to orders, to maintain discipline; self-control; ability to master problems posed in the course of duty.

III. Devotion to duty: satisfactory accomplishment of assigned tasks; steadiness and perseverance; desire to carry out assignments to maximum of ability; interest in and liking for duties; motivation toward self-improvement with respect to the job; care in the maintenance and checking of equipment; concern for safety precautions; attention to details.

IV. Sociability: ability to interest others and to achieve acceptance by them; openness and frankness in social relations; concern for the needs, welfare, and proper instruction of others; helpfulness; "esprit de corps"; liking for others; ability to correct own mistakes and to exert self-control in the interest of being agreeable to others; sense of responsibility toward subordinates.

V. Keenness and vitality: liking for hard work; confidence; zest; assertiveness; desire to attract attention; willingness to run risks; ability to impart enthusiasm to others.

VI. Initiative: liking for action; self-reliance; foresight; planning ability; executive and administrative ability.

VII. Physical excellence: good physizue; good health; toughness; calmness of temperament; suitability for combat training; fondness for sports.

VIII. General ability: above-average intelligence; adequate education.

IX. Practical intelligence: clear-signtedness; common sense; foresight; serious-mindedness; maturity.

X. Prior special training: possession of knowledge and experience deriving from practice of a trade or from special training; competence in practical matters.

XI. Manual skill: cleverness with hands; good judgment in routine activities.

XII. Knowledge of techniques and tactics of branch to which assigned: knowledge of branch regulations and instructions; familiarity with the techniques and tactics of the branch; knowledge of the use and operation of weapons.

Summary

Faced with the problem of selecting large numbers of non-commissioned officers for a rapidly expanding army, the author and his colleagues surveyed field officers and training school commanders to obtain data on NCO job requirements. The school commanders were found to emphasize intellectual and physical qualities, whereas the unit commanders stressed practical skills and dexterity in the handling of men. These differences in the criteria of success were attributed to the operation of selective factors sharpened by differences in viewpoint, and a new investigation was launched in order to identify more precisely the qualities related to success on the job.

The procedure of the new investigation combined the method of group discussion with use of the critical incident technique. A series of eleven group discussions (one for the armored division, and two each for the artillery, infantry, service and transport, engineering and administrative divisions) was held in 1954 to tap the opinions of superior officers regarding the distinguishing characteristics of successful NCO's. Each two-session discussion was attended by eight unit commanders from the various service divisions specified.

The participants were asked during the morning session to describe the specific behaviors manifested in defined situations by particularly good and bad NCO's, and by individuals who could be expected to develop into NCO's of either type. It was requested that all remarks be based on personal observation. The afternoon sessions were devoted to open discussions during which the group members described the chief overall qualities they believed to characterize the successful NCO, and rated these traits for relative importance. Three psychologists and the author acted as observers and recorded the proceedings at each session.

The extensive sample of behaviors obtained from the discussions was reduced by classification to a list of 50 items. This list was subsequently sent to 14 school commanders, 17 training-center commanders, and 265 unit commanders for evaluation. The raters were asked to assign 5, 10, 20, 10, and 5 items, respectively, to five predefined classes arranged in decreasing order of importance.

Cross-comparison of the obtained ratings indicated that the NCO requirements for various branches of the service are so different that a generalized picture of the good NCO is difficult to draw. In order to throw light on the specific nature of inter-branch differences, a further analysis was undertaken.

To facilitate treatment of the data, it was decided to eliminate items of low or debatable importance from the list, and to group the remaining behaviors into psychologically coherent clusters. Accordingly, items were selected which had generally been considered to be of "average," "great," or "very great" importance by all raters who had taken part in the study (including the group discussion participants). Items from the lowest two categories of importance, or about which the raters had generally failed to agree, were discarded. The behaviors retained were then grouped subjectively into 12 clusters, equal importance and independence of items being assumed, and multiple item placement being permitted. These clusters were entitled, respectively, "Motivation for Duty," "Ascendancy," "Devotion to Duty," "Sociability," "Keenness and Vitality," "Initiative," "Physical Excellence," "General Ability," "Practical Intelligence," "Prior Special Training," "Manual Skill," and "Knowledge of Techniques and Tactics of Branch to Which Assigned."

Each of the clusters was then assigned a weight, based on the number of items it contained, in such a way as to yield a total score across clusters of 100 points. The resulting set of 12 values was considered to represent a "basic reference pattern" applicable to NCO's in general. To obtain specific patterns for each service division, the items contained in the clusters were further classified into three groups according to the degree of inter-rater agreement regarding the importance of the behaviors in question. Items in the three groups, (arranged in decreasing order of degree of agreement), were assigned weights of 4, 3, and 2, respectively.

Weighted sums for each of the 12 previously derived clusters were then calculated for each service branch, the totals across clusters again being set at 100. The result was a table containing various sets of values considered to represent the specific patterns of NCO qualities typically required, in greater or lesser degree, by each service branch (see Table 1 for excerpt). By comparing the specific patterns with the basic reference pattern, estimates could also be made of the relative degree of importance assigned by each branch to the various traits represented by the clusters.

In general, the results of the research indicated that, due to differences in branch requirements, it is virtually impossible to formulate a general picture of an NCO who would be equally successful in any division of the service. Although some qualities (devotion to and motivation for duty, ascendancy, sociability, initiative, practical intelligence, and specific knowledge and skills) were found to be universally required, the degree of importance assigned to them varied from one branch to another. Finally, some traits, e.g., high general intelligence, special prior training, vitality and physical excellence, appeared to be specific to particular branches. The findings, therefore, suggest that the selection of NCO's should be conceived in differential terms.

A section detailing the content of the 12 item clusters is appended.

Résumé

Se trouvant confrontés avec la nécessité de sélectionner des sous-officiers en grand nombre pour une armee rapidement croissante, l'auteur et ses collaborateurs ont interrogé des officiers de l'active ainsi que des commandants de centres d'entraînement afin de se documenter sur les impératifs inhérents à la tâche du sous-officier. Il en appert que les commandants de centres d'entraînement insistent sur les qualités intellectuelles et physiques, tandis que les commandants d'unités de l'active soulignent les capacités pratiques et l'habileté à manier les hommes. Ces divergences quant aux critères de succès ont été attribuées au jeu de facteurs sélectifs accentué par des différences de point de vue; aussi une nouvelle enquête a-t-elle été menée afin de permettre de déterminer avec plus de précision les qualités du bon sous-officier.

La nouvelle enquête a combiné la méthode de la discussion en groupe avec le recours à la Technique de l'Incident Critique. Une série de onze discussions en groupe (une pour la division blindée et deux pour chacune des divisions: artillerie,

infanterie, services et transports, génie, administration) a été organisée en 1954 dans le but de recueillir les opinions d'officiers supérieurs quant aux caractéristiques déterminant le succès du sous-officier dans sa mission. A chacune de ces discussions, en deux séances, assistaient huit commandants d'unités relevant des diverses divisions précisées ci-dessus.

Au cours des séances du matin, les participants étaient invités à décrire le comportement caractéristique dans des situations bien définies de sous-officiers particulièrement bons et d'autres particulièrement mauvais, ainsi que d'individus susceptibles de se développer ultérieurement soit en l'une des directions soit en l'autre. Toutes les remarques devaient être basées sur des observations personnelles. Les séances de l'après-midi étaient consacrées à la ouverte discussion, au cours de laquelle les membres des groupes précisaient les principales qualités générales qu'ils estimaient caractériser le sous-officier souhaitable, et classaient ces qualités par ordre d'importance relative. Trois psychologues de concert avec l'auteur servaient d'observateurs et tenaient le procès-verbal des travaux de chaque séance.

Le grand nombre des traits de comportement dégagé des discussions fut réduit, par voie de classification, à une liste de 50. Cette liste fut ensuite soumise pour appréciation à 14 commandants d'écoles, 17 commandants de centres d'entraînement et 265 commandants d'unités. Les commandants étaient invités à ranger 5, 10, 20, 10, et 5 comportements respectivement dans cinq classes pré-établies ordonnées par importance décroissante.

La comparaison des classements a fait apparaître de telles différences dans les qualités requises selon les divers services qu'il est difficile d'en tirer un portrait théorique universellement valable du parfait sous-officier. Une nouvelle enquête a donc été entreprise pour permettre de tirer au clair la nature spécifique des différences entre les divers services.

Pour faciliter le maniement des données, il fut décidé d'éliminer de la liste les éléments d'importance secondaire ou discutable, et de grouper les comportements restants en catégories présentant une certaine cohérence psychologique. En conséquence, seules les qualités qui avaient été généralement considérées de "moyenne," "grande" ou "très grande" importance par tous les appréciateurs ayant participé a l'étude (y compris les participants aux discussions en groupe) furent retenues. Les éléments des deux dernières catégories d'importance ou qui n'avaient pu rallier l'accord général des appréciateurs, furent écartés. Les comportements retenus furent alors groupés en 12 catégories, l'égalité d'importance et l'indépendance mutuelle des éléments étant supposées, et des classements multiples étant permis. Ces catégories furent respectivement intitulées "Motivation quant à la Mission," "Ascendant," "Dévouement dans la Mission," "Sociabilité," "Enthousiasme et Vitalité," "Initiative," "Excellence Physique," "Aptitude Générale," "Intelligence Pratique," "Formation Spéciale Préalable," "Dextérité Manuelle" et "Connaissance des Techniques et Tactiques Particulières aux Services Déterminés."

Chacune des catégories fut alors affectée d'un coefficient, fonction du nombre d'éléments qu'elle comportait, de façon à donner, pour l'ensemble des coefficients, un total de 100 points. La série résultante de 12 valeurs fut considerée comme un "profil fondamental de référence" applicable aux sous-officiers en général. Pour en tirer des profils spécifiques pour chaque service, on a reclassé les traits en trois groupes selon le degré d'accord des appréciateurs quant à leur importance. Les comportements des trois groupes (ordonnés par degré de concordance décroissant) furent affectés des coefficients 4, 3, et 2 respectivement.

Les sommes des coefficients pour chacune des 12 catégories déjà constituées furent alors calculées pour chaque service. Le résultat en est un tableau donnant divers jeux de valeurs considérés comme profils représentatifs de qualités typiquement requises des sous-officiers, à un degré plus ou moins élevé, par chaque service (voir la Table 1, qui en donne un extrait). Une comparaison de ces profils spécifiques avec le profil fondamental de référence a également permis d'établir des estimations de l'importance relative attachée dans chaque service aux diverses qualités représentées par les catégories.

En général, les résultats de ces recherches ont montré que, en raison des différences entre les exigences des divers services, il est virtuellement impossible d'élaborer un portrait théorique général du sous-officier susceptible de réussir avec une chance égale dans toutes les branches du service. Bien que certaines qualités (dévouement et motivation quant à la mission, ascendant, sociabilité, initiative, intelligence pratique, connaissances et habiletés specifiques) apparussent universellement requises, leur importance variait d'un service à l'autre. Enfin, certains traits (par exemple intelligence générale élevée, formation spéciale préalable, vitalité, excellence physique) ont paru propres à certains services. Les constatations donc suggèrent que la sélection de sous-officiers doive procéder d'une certaine différenciation.

Une annexe donne le contenu des 12 catégories de traits.

Commentary by Discussant

Prof. Raymond G. Bonnardel (France):*

The communication of Prof. Delys is especially interesting because, up to the present, there have been few quantitative studies dealing with the non-commissioned officer. This is rather difficult to understand when one considers the large amount of research on senior officers, whose position is psychologically less interesting than that of the NCO. Standing as he does between the officers and the troops, obeying on the one hand and commanding on the other, the NCO plays a role very similar to that of the industrial foreman, whose position between labor and management has been the focus of considerable attention.

A distinguishing feature of Mr. Delys' investigation is that, unlike previous studies, it deals not solely with infantry NCO's, but with a variety of arms and services. By so doing, it throws into relief the differences in requirements which exist among the various branches.

It may be of interest to compare the qualities identified in the present research with the traits found to be characteristic of the successful NCO by other authors. As just mentioned, most of these studies have dealt with infantry NCO's, so comparison with the findings of Mr. Delys can be carried out only with respect to that branch.

*Translated from the French.

In the present study, it was found that the four most valued characteristics of the infantry NCO are, in decreasing order of importance, those of ascendance, sociability, vitality, and motivation for duty. By comparison, a study carried out in England by Valentine, and published in his book, "The Human of most importance: leadership and the ability to command, love of work, and general intelligence, in that order.

It will be seen that Valentine's first quality, leadership, compares well with that of Mr. Delys: ascendance. Similarly, love of work, Valentine's second requirement, corresponds well with Delys' third cluster: keenness and vitality. With respect to intellectual qualities, however, there is some divergence between the findings of Valentine and those of Delys. The same is true of sociability, which was the second most heavily weighted trait in the infantry pattern obtained by Delys, but was classified as least important by Valentine.

As regards the question of social qualities, I am personally inclined to find the results of Mr. Delys the more intellectually compelling. It should be noted, however, that in their studies reported in "The American Soldier," Stouffer and his colleagues found almost exactly the same thing as Valentine with respect to the relative unimportance of sociability as a characteristic of the NCO. Interestingly enough, in the study of Stouffer (as in the investigation of Delys) the trait ratings were obtained from the troops. But, this is not all. Stouffer also carried out a study based on the opinions of soldiers, and in this he obtained results similar to those of Delys, i.e., sociability was considered to be one of the most essential traits of the NCO.

Thus, two investigations, one based on the opinions of soldiers (Valentine) and one based on the opinions of officers (Stouffer), agree as to the comparative unimportance of the social qualities of the NCO, and two further studies, again based, respectively, on the opinions of soldiers (Stouffer) and the opinions of officers (Delys), agree as to the relative importance of sociability as a quality of the NCO. Put otherwise, it can be said that English soldiers and American officers do not seem to value sociability particularly in the infantry NCO, whereas American soldiers and Belgian officers regard it rather highly. Cultural differences may be partly responsible for these discrepancies. It should also be noted that the studies mentioned were carried out some time ago, and it may be that a certain evolution has since taken place in the opinions of officers. We find today a similar phenomenon in industry: management is beginning to place more and more emphasis on social qualities in its choice of foremen.

As regards the difference between the findings of Delys and Valentine with respect to intellectual qualities, two points are relevant. First, as Prof. Delys has pointed out, selection of NCO's for general intelligence is carried out in Belgium prior to the candidates' entrance into training school. Second, the results of his initial survey showed that the unit commanders think the schools place too much emphasis on intellectual qualities and not enough on human and professional education. For both these reasons the unit commanders may have tended to underestimate the importance of intelligence in the NCO. To draw a parallel again with industry, we find there that general ability is a rather important characteristic of the successful foreman. When intellectually able men are chosen as foremen, it is much easier to train them to recognize and remedy incorrect behavior on the part of their workers.

It might, incidentally, have been interesting if Prof. Delys had attempted to establish two different sets of patterns, viz., patterns derived on the one hand from the opinions of the school and training-center officers, and on the other from the opinions of the unit commanders. Such a study might have yielded further insight into the nature of the differences between their respective conceptions of the good NCO.

In conclusion, I should like to ask the author to what extent it has been possible to apply the findings of his study to practical problems of selection, guidance, and training in the Belgian army.

Reply

Prof. Louis Delys:*

I should like to thank Prof. Bonnardel for his interesting and helpful comments, and attempt to reply to the questions he has raised.

With respect to the difference between our findings and those of others regarding the importance of sociability in the NCO, it may very well be that a certain evolution in the attitudes of officers has taken place, reflecting a general de-emphasis on authoritarianism in the culture and in the army. It can be noted in this connection that the participants in the group discussions (the unit commanders who provided the original list of behaviors, and whose opinions were taken into account in composing the patterns) were selected and sent to us by the authorities. They sent us the best to be had. Thus, we may have here a kind of indirect support for Prof. Bonnardel's hypothesis. It is assumed that individuals tend to emphasize the desirability of those qualities possessed by themselves, and given that the group discussion members were successful officers, then it may be that the commanders who rated the list of behaviors agreed as to the value of a "social sense" in the NCO because in the modern army this quality of the officer has actually taken on an increased importance.

As regards the utilization of the findings, they have not been applied in a completely systematic manner. However, we always emphasize to those responsible for selection the above point about the NCO's social capacities. Thus, at present, the personnel officers take into account both general ability and social qualities in selecting candidates for NCO training.

With respect to the matter of general intelligence, it seems to me that although the comments of Prof. Bonnardel are most cogent, one should not make the mistake of over-emphasizing the importance of intelligence in the subordinate officer. Naturally, he must possess sufficient general ability to carry out his duties effectively. On the other hand, there tends to be a curvilinear relationship between intelligence and success in relatively undemanding jobs such that the individual of superior intellect may become bored and dissatisfied with low level work, and ineffective in carrying out tasks which for him present no challenge.

*Translated from the French.

Finally, I agree with Prof. Bonnardel that it would have been of interest to derive separate patterns based, respectively, on the opinions of training school and unit commanders. However, it was considered inadvisable to do this because of the comparatively small number of training school commanders available to us. We had only 32 training school and training center officers serving as raters, whereas the number of unit commanders totaled 265. It was decided, therefore, to combine the date for purposes of item selection and weighting in order to maximize the reliability of the obtained patterns.

THE ATTITUDE OF MILITARY AUTHORITIES TOWARD SCIENTIFIC PSYCHOLOGY

Joseph M. Hamon
School of Application, Military Health Service
Paris, France

In a certain sense, psychology and sociology have always at least in an informal manner, constituted an integral part of the government and handling of men. But, in this age of rapid technological advance, mass communications, and collective tensions, the knowledge to be gained from rigorous application of psychological and sociological analysis has tekan on a particular significance.

The army, as a social organization responsible for the moding of its members into an efficient fighting force, can draw much profit from the understanding of behavior to be gained from social scientific research. The attitudes of its governing officials toward psychological science is, therefore, a matter of major significance.

This paper concerns the nature of armies as psycho-social entities, and the effects of their characteristic goals and modes of operation on the attitudes toward scientific psychology of those in positions of authority.

The Army as a Social Institution

Today, it is recognized that military organizations do not exist in a sociological vacuum, and that, like other institutions, they are affected by the changes and pressures arising in the surrounding culture. Moreover, the principles governing the internal functioning of military communities can be likened in many respects to those ruling other types of social organizations.

The fighting value of troops can be said to depend not only on their internal cohesion, but also, among other factors, on the adaptation of individual men to unfamiliar living conditions, in groups governed by characteristic rules and regulations. Armies no longer consist of professional soldiers but of conscripts, and the importance of this fundamental fact should not be overlooked. At present, it is the fate of young men to find themselves abruptly uprooted from their social milieux and detached from their trades and professions to enter a military organization which imposes on them new duties and a new routine of life. The ease and speed with which they adapt, or can be helped to adapt; to these suddenly changed conditions is a matter of concern not only to the individuals involved but also to the army itself. Thus, the military authorities will necessarily give serious attention to whatever scientific psychology can contribute to this and other problems of military efficiency.

In many respects, the difficulties encountered by the new conscript are no different from those he might meet in any section of society having graded ranks and stations. What distinguishes other social institutions from the army, however,

is that in time of peace the army must be prepared to make war, and the controls and training it imposes are directed toward that end. There can be no doubt but that this special purpose of the army specifically conditions both its functioning and its structure. In time of war, the entire military body must be capable of operating as a unit, and the behavior of its individual members must conform to standardized rules of action. It is for this reason that the structure of the army must be strongly hierarchical, and that its authorities must place heavy emphasis on discipline and on the control of individual initiative.

It can be seen that the special mission of the army may render questionable the application to military situations of principles drawn from the study of other social groups. It is in relation to the fighting man himself, studied within his military environment, that relevant psycho-sociological investigations must be carried out. Such investigations, however, will meet with various difficulties of which it is necessary to take stock, and for which it is necessary to be prepared. Among these problems is that of resistance on the part of military authorities to research suggestions which may be perceived as constituting attempts at "foreign determinism."

Factors Conditioning Military Attitudes toward Scientific Psychology

Any social group orienting its communal life toward the achievement of certain goals tends to be sustained in its organization and animated in its activities by certain fundamental beliefs, whose strength depends on the length of time the group has been in existence. Such a group naturally tends to resist outside efforts to change its structure and traditions. Furthermore, attempts to induce changes may bring to the surface latent problems, cause vague feelings of uneasiness in the group, and, as a result, hinder without any proven justification the actions of those who hold and exercise power. In the military, those officials are responsible for protecting and safeguarding the entire civic community. This mission is of such great importance that any influences perceived as interfering with its accomplishment are liable to meet with especially strong resistance.

It would perhaps be better to speak here of "reserve" rather than of "resistance." By this is meant a reluctance to accept ideas which are not in accord with accepted practice, and have not yet been subjected to the test of time and experience. All such reserves, with their attendant appeals to tradition and experience, all defensive reactions which may appear to be unduly conservative, merely represent symptoms of a desire to prevent internal disorder and to preserve military effectiveness. Hence, to be rapidly accepted, suggestions for change and improvement resulting from research must be perceived as contributing directly to the accomplishment of the ends to which the army is dedicated.

Guiding Principles for the Application of Scientific Psychology to Military Problems

In the preceding discussion, the distinguishing psycho-social characteristics of the military organization have been examined, the special nature of its institutional role and purpose has been emphasized, and how these and other factors condition its structure, its functioning, and the attitudes of its authorities toward external influences has been pointed out. What are the conclusions to be drawn from this analysis with respect to the problem of how best to apply the techniques of scientific psychology within a military context?

First, it can be concluded that in the planning of investigations attention should be given both to the unique features of the military institution and to the general attitudes held by the army community and its governing officials. All such studies should be guided in part by the military authorities, and should take into consideration suggestions drawn from their experience. Planning of this nature should help to broaden the field of application of the obtained findings.

Second, since an army forms an integral part of its nation's culture, and is continually affecting and being affected by it, specific problems should be attacked within the framework of a broad psycho-sociological perspective.

Finally, the purpose and reison d'etre of the military organization—its responsibility for the defense and protection of the civic community—should never be lost from sight. Studies should be so oriented as to contribute ultimately in some way to the improvement of military efficiency and combat effectiveness. Among specific research problems which can be mentioned here are those of personnel selection, adaptation of the individual to military life, acceptance of discipline, mental health, and morale. In such studies, the potential contribution of those in closest contact with the troops—the unit commander and medical officer—should not be overlooked.

In conclusion, it can be reiterated that the military authorities will necessarily give serious consideration to any findings or principles which will help them to perfect the instrument they are called upon to handle. The need for such information is of particular importance in this age of rapidly accelerating technical progress. The social scientist, if he is to contribute effectively to this endeavor, must keep ever in mind the mission with which the military authorities are charged, and orient his planning and research in terms of that obligation.

Summary

Since scientific psychology can aid the army in its function of protecting the civic community, the attitude of military officials toward psychological research is an issue of major significance. The present paper deals with the distinguishing psycho-social characteristics of military organizations, the effects of these characteristics on the attitudes of those in positions of authority, and the implications which result for the application of psychology to military affairs.

A military organization is affected by changes in the surrounding culture, and shares many functional characteristics in common with other social institutions. However, an army is distinguished from other social groups by reason of its partirular mission—that of defending the civilian society. This special purpose of the army necessarily conditions its structure and functioning, e.g., strongly hierarchical organization, emphasis on discipline, and the control of individual initiative, in ways which may render questionable the extrapolation to military situations of principles drawn from the study of other social groups. Needed psychological research must therefore be carried out within the military organization itself. Such research will meet with various problems, of which one of particular importance is resistance on

594869 O - 61 - 7

the part of army authorities to external influences perceived as interfering with military efficiency.

Any social group, organized for the purpose of achieving certain goals, possesses characteristic beliefs and traditions which it will resist outside efforts to change. Furthermore, attempts to induce such changes may hinder the group leaders in their efforts to implement the group mission. In the case of the army, this mission is of such great importance that influences regarded as capable of provoking internal disorder or otherwise interfering with military effectiveness will meet with especially strong resistance. Thus, if they are to be accepted, suggestions for change emanating from psychological research must be perceived by military officials as contributing directly to the accomplishment of the ends to which the army is dedicated.

Several conclusions can be drawn from these observations. First, the planning of psychological research should take into account both the unique features of the military organization and the attitudes of its governing officials. All such studies should be guided in part by the military authorities, and should give consideration to suggestions deriving from their experience.

Second, since the army forms an integral part of its nation's culture, specific problems should be attacked within the framework of a broad psycho-social perspective.

Finally, the military psychologist should never lose sight of the mission with which the army is charged, and should attempt to insure that his studies will contribute to the improvement of military efficiency and combat effectiveness. The importance to this endeavor of the information to be obtained from those in closest contact with the troops—the unit commander and the medical officer—should not go unrecognized.

In conclusion, it is emphasized that the military authorities will necessarily give serious consideration to any findings or principles of scientific psychology which will help them in the accomplishment of their purpose.

Résumé

La psychologie scientifique pouvant aider l'Armée dans sa fonction de protection de la communauté civique, l'attitude des autorités militaires à l'égard de la recherche psychologique est de la plus haute importance. La présente communication est consacrée aux caractéristiques psycho-sociales distinctives des organisations militaires, à certains effets de ces caractéristiques sur les attitudes des autorités, et à ce qu'elles impliquent quant à l'application de la psychologie aux affaires militaires.

Toute organisation militaire est affectée par les changements survenant dans le milieu culturel, et partage avec d'autres institutions sociales de nombreuses caractéristiques fonctionnelles. Toutefois, une armée se distingue des autres

groupements sociaux en raison de sa mission particulière: la défense de la société civile. Cet objet tout spécial de l'armée en conditionne nécessairement sa structure et son fonctionnement, par example, une organisation fortement hiérarchisée et l'importance accordée à la discipline et le contrôle de l'initiative individuelle, et dès lors rend problèmatique l'extrapolation aux problèmes militaires des principes dégagés des études civiles. Les recherches psychologiques nécessaires doivent donc se poursuivre au sein de l'organisation militaire même. Ces recherches se heurteront à diverses difficultés, dont l'une, particulièrement importante, est la résistance opposée par les autorités militaires aux influences vues comme entraves possibles à l'efficacité militaire.

Tout groupe social organisé en vue de la réalisation d'objectifs particuliers a ses croyances et traditions caractéristiques, et résiste aux efforts tendant à les modifier. De plus, des tentatives pour provoquer un changement peuvent gêner les chefs du groupe dans leurs efforts pour remplir la mission du groupe. Dans le cas de l'armée, cette mission revêt une telle importance que toute influence susceptible d'engendrer des désordres internes, ou d'entraver d'une façon quelconque l'efficacité militarire, suscitera une opposition particulièrement énergique. Pour être acceptée, des suggestions émanant de recherches psychologiques devront donc apparaître aux autorités militaires comme contribuant à la réalisation des objectifs de l'armée.

Les observations qui précèdent conduisent à certaines conclusions. Tout d'abord, la programmation de la recherche psychologique doit tenir compte tant du caractère spécial de l'organisation militaire que des attitudes des autorités responsables. Toutes ces études devront être guidées en partie par les autorités militaires, et prendre en considération les suggestions que l'expérience leur aurait inspirées.

En second lieu, l'armée étant partie intégrante de la culture nationale, les problèmes particuliers devront être abordés dans une perspective psycho-sociale étendue.

Enfin, le psychologue militaire ne doit jamais perdre de vue la mission impartie à l'armée, et devra s'efforcer d'assurer que ses études contribuent à améliorer l'efficacité militaire et le rendement au combat. L'importance que présente à cet égard l'information que peuvent transmettre ceux qui sont au contact le plus étroit avec les troupes (le commandant d'unité et l'officier médicin) ne doit pas être ignorée.

En conclusion, il est souligné que les autorités militaires accueilleront nécessairement avec la plus haute attention toutes découvertes ou tous principes de psychologie qui pourraient les assister dans l'accomplissement de leur mission.

Commentary by Discussant

Dr. Luigi Meschieri (Italy):*

General Hamon has presented us with some cogent observations on the nature of armies which reach to the roots of the psychological problems of military organizations. I was particularly struck by his emphasis on the opposition between the preparation of young men as constructive members of civilization and their training as soldiers for a destructive event such as war. To aid adaptation to this contrast, it would seem important to attempt to develop in young men a continuous consciousness of the value of tasks necessary to the defense and protection of their country.

As to the attitudes of military authorities toward behavioral science, we are no doubt justified in expecting that these attitudes will be favorable when scientific psychology leads to the same results as "intuitive" psychology. But what will happen in the very possible, and even frequent, case where common sense arrives at policies and programs different from those which would have been obtained from empirical research? It would perhaps also be desirable to examine the attitudes of commanding officers in relation to this critical question.

In discussing the psycho-social characteristics of armies, General Hamon pointed out that military organizations share certain features in common with other social groups. This facilitates the military application of psychological knowledge, and permits the psychologist to draw, within limits, upon the results of investigations carried out in the civic community.

In connection with this point, it can be noted that the attitudes of civilian authorities toward scientific psychology are often similar to those of military officials. For example, psychologists may receive from both requests with which they are not always prepared to cope. Thus, just as industrial executives may ask us for methods by which personality characteristics such as interest in work, good will, and company loyalty can be developed, so the military authorities may want us to suggest means for diagnosing and developing such qualities as courage, conformance to discipline, and so on. In such situations, we often find ourselves hindered by the problem of defining these behaviors in terms of specific, stable personality traits, and by the difficulty of developing satisfactory diagnostic and educational methods. In the present state of our science, we often are not able to provide solutions to all of the problems put to us.

Thus, there is a necessity for full understanding by the military authorities of the nature and extent of the contribution which psychology is capable of making. Sometimes, faced with urgent appeals and with the difficulty of explaining to those unversed in his field why they cannot be met, the psychologist is tempted to make suggestions which, in actuality, lack a firm empirical foundation. It is to be hoped, however, that, while developing his techniques to the maximum, he will refrain from too zealous penetration into fields outside the domain of his professional competence, and avoid risking damage to his own prestige and to that of his science. It is necessary that the psychologist strictly respect the limits and present capabilities of his discipline, and define them precisely for the military authorities who ask his advice. It is only within these limits that it will be possible for him to

*Translated from the French.

make a real contribution. This is another way to foster sound and favorable attitudes on the part of military officials toward scientific psychology.

Reply

Surgeon General Joseph M. Hamon:*

I should like to thank Dr. Meschieri for his comments, with which I am in general agreement. There is one point he drew to your attention which perhaps was not sufficiently emphasized in my paper. This concerns the necessity of providing a nation's youth with what we might call a moral preparation for military effort. It is essentially a social and political problem which falls more or less outside the domain of military responsibility, and which should be attacked prior to the age of entry into service. Psychologists, working in cooperation with civilian and military authorities, will no doubt be able to contribute to this important endeavor.

*Translated from the French.

ATTITUDES AND THEIR CORRELATES: A STUDY OF PILOT CANDIDATES

A. de Brisson and P. Legrand
Air Force Center of Psychological Studies and Instruction
Versailles, France

An investigation in which attitude measures were developed by means of hierarchical scaling methods for use in studying the motivations of pilot candidates is reported herein. The ultimate aim was to develop instruments which would permit analysis of the changes in attitudes which occur during the course of pilot school, and lead to suggestions for the improvement of training procedures. A more immediate purpose was to study the relations of measured attitudes to the aptitude and biographical variables comprising the existing pilot selection battery.

Background of the Research

The pilot selection battery currently in use by the French Air Force consists of an adaptation, made in 1945, of the American Aircrew Battery. In addition to written and psychomotor aptitude tests, the battery includes four homogeneous biographical inventory keys, which were constructed by applying the Loevinger-Gleser-DuBois Technique (Loevinger, Gleser, & DuBois, 1953) to data obtained from personal history questionnaires. These four keys, developed and cross-validated on large samples of French pilots, were found to contribute significantly to the prediction of success in training school. They are entitled, respectively, "Handiness at Odd Jobs" (Key A), "Practice of Certain Sports" (Key B), "Family Environment" (Key C) and "Aviation Interests" (Key D). Keys A and D, although the most valid predictors, were found to be rather highly related to the pilot aptitude stanines. Keys B and C are virtually independent of the pilot stanines, but possess considerably lower validity.

It can be seen that current pilot selection procedures yield data on criterion-relevant aptitudes, achievements, and personal history characteristics. Although motivations and attitudes can perhaps be indirectly inferred from such data, they are not directly measured thereby. It was considered desirable, then, to develop objective measures of attitude and motivational variables which may be adapted to pilot school. Because attitudes vary with sisuational changes, and can be expected to fluctuate throughout the course of training, measures made at the time of recruitment cannot be used for purposes of selection. Longitudinal studies of such changes, however, may lead to suggestions for the improvement of policies and methods. In the meantime, it is of general theoretical interest to determine how attitudes measured at the beginning of training are related to biographical and aptitude variables.

Procedure

The attitude measures with which the present report is concerned were developed by administering questionnaires containing a wide variety of items to a

large sample of pilot candidates. The obtained responses were analyzed by hierarchical scaling methods using several criteria of internal consistency. In addition to Guttman's index of reproducibility (1947), the homogeneity criteria of Loevinger (1947), and of Green (1956) were utilized. Several criteria of consistency were used in order to obtain information on how the results of methods based on different rationales compared with one another.

Five attitude dimensions emerged from this analysis, and were subsequently cross-validated on several samples of pilot candidates. The results presented in the following section are based on a group of five hundred subjects.

In order to determine how scores on the attitude scales were related to available aptitude and biographical measures, an intercorrelation matrix was computed for the five attitude scales, the four previously derived biographical inventory keys, and pilot stanines based on the aptitude tests of the selection battery. The matrix was then factor-analyzed by the Thurstone method, the resulting centroids being rotated to an orthogonal simple structure solution.

Results

Analysis of the attitude items yielded five homogeneous scales, entitled "Flight Motivation," "Military Motivation," "Fighter Pilot Motivation," "Taste for Risk," and "Sociability." Table 1 gives the number of items contained in each scale and the internal consistency coefficients obtained by the Guttman, Loevinger, and Green methods for a sample of five hundred pilot candidates. In connection with these coefficients, the values considered as the minimum required for a satisfactory scale were .90, .30, and .25 for the Guttman, Loevinger, and Green techniques, respectively.

TABLE 1
INTERNAL CONSISTENCY OF ATTITUDE SCALES AS DETERMINED ON A SAMPLE OF 500 PILOT CANDIDATES

Scale	No. of Items	Coefficients of Internal Consistency		
		Guttman	Loevinger	Green
Flight Motivation	8	.930	.489	.492
Military Motivation	8	.926	.513	.389
Fighter Pilot Motivation	5	.918	.442	.500
Taste for Risk	6	.900	.334	.274
Sociability	5	.918	.254	.193

The intercorrelation matrix, calculated prior to carrying out the factor analysis, is presented in Table 2. Analysis of this matrix yielded, after rotation, three orthogonal factors entitled "Aviation Aptitude and Motivation," "Hero Complex," and "Socio-cultural Status." The variables loading above .12 on these three factors, together with their rotated factor saturations, are presented in Table 3.

TABLE 2
INTERCORRELATION MATRIX BASED ON A SAMPLE OF 500 PILOT CANDIDATES

	1	2	3	4	5	A	B	C	D	Pilot Stanine
1		−.01	.12	.19	.03	.10	.05	.07	.49	.33
2			.10	.19	.11	.02	.05	−.09	−.05	−.16
3				.46	.01	−.07	.08	−.01	.05	−.07
4					.14	.03	.17	.12	.05	−.02
5						.07	.22	.16	.05	−.04
A							.15	.16	.16	.29
B								.26	.22	−.06
C									.20	.04
D										.34
Pilot Stanine										

Note: Coefficients of .11 or above are significant at the .01 level of confidence.

Key

Attitude Scales	Biographical Inventory Keys
1 – Flight Motivation	A – Handiness at Odd Jobs
2 – Military Motivation	B – Practice of Certain Sports
3 – Fighter Pilot Motivation	C – Family Environment
4 – Taste for Risk	D – Aviation Interests
5 – Sociability	

TABLE 3
CENTROID FACTORS OBTAINED FROM DATA IN TABLE 2
(After Rotation to Simple Structure)

	Factors	Constituent Variables	Rotated Loadings
I	Aviation Aptitude and Motivation	Pilot Stanine	.66
		Aviation Interests (D)	.61
		Flight Motivation (1)	.60
		Handiness at Odd Jobs (A)	.34
II	Hero Complex	Taste for Risk (4)	.73
		Fighter Pilot Motivation (3)	.63
		Military Motivation (2)	.24
		Flight Motivation (1)	.22
III	Socio-cultural Status	Family Environment (C)	.55
		Practice of Certain Sports (B)	.51
		Sociability (5)	.37
		Aviation Interests (D)	.30
		Handiness at Odd Jobs (A)	.30
		Taste for Risk (4)	.22

Note: Letters and numerals appearing in parentheses after titles of variables identify, respectively, biographical inventory keys and attitude scales.

Discussion

The attitude scales

Study of Table 2 shows that, in general, the five attitudes scales are relatively independent of one another. An exception is the comparatively high correlation between "Fighter Pilot Motivation" and "Taste for Risk," a finding which is not difficult to rationalize.

The complete lack of relationship between the "Flight Motivation" and "Military Motivation" scales is a phenomenon of some interest. It would appear that, as a group, individuals interest in flying are not characteristically motivated to pursue such a career within a military context.

Although "Flight Motivation" bears a fairly high positive relation to the pilot stanines derived from the aptitude tests of the selection battery, the correlation between "Military Motivation" and the pilot stanines is significantly negative. Thus, it is found that interest in a military career is unrelated to flying motivation, and that the pilot candidates interested in a military career are those who tend on the average to be less gifted from the standpoint of flying aptitude. This is unfortunate, since the considerable discrepancy between military salaries and those offered in civilian aviation already creates problems as to how the service is to retain its able flight personnel. To find, in addition, that it is the less gifted pilot candidates who have the greatest interest in a service career is indeed discouraging. One solution to the problem of how to keep the most capable pilots in the service might be to extend the length of military contracts.

It can also be noted that the five attitude scales tended to be independent of the biographical inventory keys. Understandable exceptions were the relatively high correlation of .49 between the "Flight Motivation" scale and the "Aviation Interests" key, and the correlation of .22 between the "Sociability" scale and the key, "Practice of Certain Sports." In general, however, the attitude scales and biographical keys seemed to be tapping different sources of variance.

A final point of incidental interest is that in the "Flight Motivation" scale the first two items deal with previous flying experience. This is the only case in the study where objective items of a biographical nature were found to be highly enough related to attitudinal items to warrant their inclusion in a scale. This finding reinforces the results of other studies which have demonstrated the importance of previous flying experience as a predictor of aviabion interests and motivation.

Factor-analytic results

The results presented in Table 3 show, first of all, that interest in and aptitude for flying tend, not unexpectedly, to hang together. Individuals highly motivated to fly also tend to be those most capable of doing so, and vice versa. Factor I obviously is related to success in training.

Factor III, "Socio-cultural Status," brings together several of the biographi- and attitude variables. All four of the biographical inventory keys and two of the attitude scales load on this factor. The significantly correlated pilot stanine and "Flight Motivation" variables have very low loadings, a fact which helps to explain the relatively low saturations of biographical inventory Keys A and D (which are

correlated with both of the above measures). The highly loaded "Family Environment" and "Practice of Certain Sports" keys were previously found, it will be remembered, to have negligible correlations both with the pilot stanines and with measures of success in pilot school. This factor, therefore, seems to be measuring sources of variance which are not associated with success in training.

The most interesting result of the factor analysis was the subsequent finding of a significant negative correlation between scores on the "Hero Complex" factor (which is comprised exclusively of attitude scales) and the variables predictive of success in training. This suggests that the recklessness and daring commonly associated with the popular stereotype of the pilot are actually contra-indicators of flying proficiency. In view of the fact that the "Flight Motivation" scale loads lightly but significantly on Factor II, it might be advisable to attempt to separate that portion of the variance of this scale associated with aviation interests and aptitudes from that portion apparently due to personality characteristics, which may lead to carelessness and the tendency to take unnecessary risks.

Summary and Conclusions

Five attitude scales were constructed by hierarchical methods of analysis for the purpose of studying the motivations of pilot candidates in military training. The intercorrelations among these scales and the biographical keys and aptitude stanines of the existing pilot selection battery were computed, and a centroid factor analysis of the obtained matrix was carried out.

The most interesting results of the study concern the lack of relation between the "Flight Motivation" and "Military Motivation" scales, and the finding of a significant negative correlation between "Military Motivation" and flying aptitude as measured by pilot stanine scores. These results underline the importance of developing means for increasing the interest of able pilot candidates in a military career.

Three orthogonal dimensions were obtained after rotation to simple structure of the factor-analytic results. Of these, the factor "Hero Complex" (which reflects recklessness and a seeking for prestige) was found to be negatively correlated with predictors of success in pilot school. The fact that the "Flight Motivation" scale loads on this factor suggests that attempts should be made to distinguish between individuals whose high scores on this scale reflect flying interests and aptitudes, and individuals who score high because of personality attributes which may lead to recklessness in flight.

In conclusion, it should be emphasized that the measurement of attitudes and motivation at the time of recruitment is conceived as only a first step in an on-going attitude testing program. It is planned, also, to use the scales developed in this study to investigate attitudinal changes during the course of training. It is expected that the results will lead to practical suggestions for the improvement of training methods and personnel policies.

References

Green, B. F. A method of scalogram analysis using summary statistics. Psychometrika, 1956, 21, 79-88.

Guttman, L. The Cornell technique for scale and intensity analysis. Educ. psychol. Measmt., 1947, 7, 247-279.

Loevinger, Jane. A systematic approach to the construction and evaluation of tests of ability. Psychol. Monogr., 1947, 61, No. 4 (Whole No. 285).

Loevinger, Jane, Gleser, Golding C., & DuBois, P. H. Maximizing the discriminating power of a multiple-score test. Psychometrika, 1953, 18, 309-317.

Summary

The report describes the results of a study in which: (a) attitude scales were developed for the future purpose of studying motivational changes which occur during the course of military pilot training, and (b) the relations of these scales to the aptitude and biographical measures of the French pilot selection battery were investigated.

The attitude measures were constructed by administering questionnaires of varied item content to a large sample of pilot candidates and analyzing the responses by hierarchical scaling methods. Five homogeneous scales emerged from this analysis: "Flight Motivation" (8 items), "Military Motivation" (8 items), "Fighter Pilot Motivation" (5 items), "Taste for Risk" (6 items), and "Sociability" (5 items). The scales were cross-validated on several samples of pilot candidates; data on internal consistancy indices computed by the Guttman, Loevinger, and Green methods for a sample of five hundred subjects are presented in Table 1. With the exception of the relation between "Fighter Pilot Motivation" and "Taste for Risk," the scales were found to be relatively independent of each other (see Table 2).

The existing pilot selection battery of the French Air Force consists of an adaptation of the American Aircrew Battery. It contains the usual written and psychomotor aptitude tests, plus four homogeneous, biographical inventory keys developed on French samples by applying the Loevinger-Gleser-DuBois technique to data obtained from personal history questionnaires. To study the relations between the newly developed attitude scales and the variables of the pilot selection battery, an intercorrelation matrix, containing the attitude scales, the biographical inventory keys, and pilot stanines derived from the aptitude tests, was computed using data from a sample of five hundred pilot candidates (Table 2). This matrix was then factor-analyzed by the Thurstone method. The analysis yielded, after rotation to simple structure, three orthogonal factors entitled: "Aviation Aptitude and Motivation," "Hero Complex," and "Socio-cultural Status." The variables loading above .12 on the rotated factors, together with their factor saturations, are given in Table 3.

The following points are emphasized:

1. The complete lack of relation between the "Flight Motivation" and "Military Motivation" attitude scales suggests that candidates highly motivated to fly are not characteristically interested in pursuing such a career within a military context, while those interested in a military career are not, as a group, highly motivated toward flight. In addition, the significantly negative correlation between "Military Motivation" and flying aptitude, as measured by the pilot stanines, suggests that it is the less gifted pilot candidates who tend to be most interested in military careers. These findings emphasize the importance of developing means whereby the military motivation of successful pilots can be increased, and the service enabled to retain its best flight personnel.

2. It is noted that, in general, the attitude scales and biographical inventory keys appear to be tapping different sources of variance. The fact that two biographical items relating to civilian flight experience appear in the "Flight Motivation" scale is consistent with other research demonstrating the importance of previous flying experience as a predictor of interest in aviation.

3. The results of the factor analysis show that interest in and aptitude for flying tend to be interrelated (Factor I). This first factor is related to pilot success. Factor III appears to be measuring socio-cultural background variance which is not associated with success in training. The most interesting result was the finding of a significant negative correlation between scores on Factor II ("Hero Comples") and predictors of success in pilot school. This indicates that the recklessness and daring popularly associated with success as a pilot are actually contra-indicators of flying proficiency and should be guarded against in selection.

In conclusion, the authors point out that attitudes and motivations should be measured periodically throughout the course of training. It is planned to use the newly developed attitude scales in future longitudinal studies, the results of which are expected to lead to practical suggestions for the improvement of training methods and personnel policies.

Résumé

Le rapport décrit les résultats d'une étude dans laquelle (a) des échelles d'attitudes ont été mises au point en vue d'étudier ultérieurement les changements de motivation qui surviennent au cours de l'entraînement du pilote militaire et (b) les relations de ces échelles avec les tests d'aptitude et les données biographiques de la batterie française de sélection de pilotes ont été analysées.

Les mesures d'attitudes ont été établies en analysant, au moyen de méthodes d'échelonnage hiérarchiques, les réponses d'un grand échantillon de candidats pilotes à un questionnaire comportant des éléments divers. Cette analyse a fait ressortir cinq échelles homogènes: "Motivation pour le vol" (8 items), "Motivation militaire" (8 items), "Motivation pour la chasse" (5 items), "Goût du risque" (6 items) et "Sociabilité" (5 items). Les échelles ont été verifiees sur plusieurs échantillonnages de candidats pilotes; la Table 1 présente les indices de cohérence

interne calculés par les méthodes de Guttman, de Loevinger et de Green pour un groupe de cinq cents sujets. Réserve faite de la relation entre "Motivation pour la chasse" et "Goût du risque," les échelles sont apparues relativement indépendantes les unes des autres (voir Table 2).

La batterie de sélection de pilotes appliquée par la Force Aérienne française consiste en une adaptation de l'American Aircrew Battery. Elle comporte les tests d'aptitude écrits et psychomoteurs, plus quatre clés homogènes d'inventaire biographique établies sur des échantillonnages français au moyen de la technique Loevinger-Gleser-DuBois. Une matrice d'intercorrélations, comportant les échelles d'attitudes, les clés biographiques et le stanine-pilote basé sur les tests d'aptitude, a été calculée à partir des données recueillies pour un groupe de cinq cents candidats pilotes aux fins d'étudier les relations entre les échelles d'attitude récemment mises au point et les variables de la batterie de sélection de pilotes (Table 2). Cette matrice a alors été analysée par la méthode factorielle de Thurstone. L'analyse a donné, après rotation a structure simple, trois facteurs orthogonaux, à sovoir: "Aptitude et motivation aéronautique," "Complexe du héros," et "Rang socio-culturel." La table 3 donne les saturations factorielles des variables chiffrant plus de 0, 12 sur ces facteurs.

Les points suivants sont soulignés:

1. L'absence totale de relation entre les échelles de "Motivation au vol" et de "Motivation militaire" indique que en tant que groupe, les candidats fortement motivés au vol ne s'intéressent pas à poursuivre une telle carrière dans un contexte militaire, tandis que ceux qu'intéresse une carrière militaire ne sont pas fortement motivés au vol. En outre, la corrélation négative significative entre "Motivation militaire" et aptitude pour le vol telle que mesurée par le stanine-pilote invite à croire que ce sont les candidats pilotes les moins doués qui ont le plus tendance à s'intéresser aux carrières militaires. Ces constatations mettent en relief l'importance d'une mise au point de moyens qui permettraient d'augmenter la motivation militaire de bons pilotes, et de conserver à la Force Aérienne les meilleurs éléments de son personnel volant.

2. Il est noté que les échelles d'attitude et les clés d'inventaire biographique semblent mesurer les sources différentes de variance. Le fait que deux items biographiques relevant de l'expérience du vol civil apparaissent dans l'échelle de "Motivation au vol" concorde avec les résultats d'autres recherches qui ont démontré l'importance de l'expérience du vol comme indice de l'intérêt aéronautique.

3. Les résultats de l'analyse factorielle montrent que la propension et l'aptitude au vol tendent a être reliées (Facteur I). Le premier facteur est impliqué dans le succès en pilotage. Le facteur III apparaît comme mesure de la variance socio-culturelle non associée a la réussite dans l'entraînement. Le résultat le plus intéressant est la constatation d'une corrélation négatave significative entre les cotes pour le Facteur II ("Complexe du héros") et des variables liées au succès à l'école de pilotage. La leçon à en tirer est que l'insouciance et la témérité que l'esprit populaire associe au succès comme pilote constituent en réalité des contre-indications d'excellence au vol et doivent éveiller la méfiance dans la sélection.

En conclusion, les auteurs maintiennent qu'attitudes et motivations doivent être périodiquement mesurées pendant toute la durée de l'entraînement. On envisagerait d'appliquer les echelles d'attitude récomment mises au point dans de futures études longitudinales, dont on attend des suggestions pratiques quant aux méthodes d'entraînement et aux lignes de conduite à suivre en matière de personnel.

Commentary by Discussant

Dr. J. L. Pinillos (Spain):

I am pleased to have the privilege of commenting on the interesting paper of Messrs. de Brisson and Legrand. In the early days of aviation psychology, Guilford and his colleagues conducted many studies aimed at identifying the personality and biographical correlates of flying success. In that research, very few significant correlations with personality variables were obtained, but, as in the work of the present authors, it was found that empirically constructed biographical inventory keys contributed significantly to prediction when added to the pilot selection battery. Dr. de Brisson and his colleague have made an additional contribution to this field by using scaling techniques and homogeneous keying methods to develop attitude and biographical measures of high internal consistency.

In Spain, we generally have been concerned less with internal consistency than with the external validity of individual biographical items, but, I must say that our efforts in this direction have not met with outstanding success. Often, we find that those items which serve to predict flying success are precisely those which correlate with our pilot aptitude measures, and do not add to the validity of the selection battery. Examples of such items are socio-economic status, and questions reflecting general information and aviation information. For items such as father's occupation and previous training of the candidates, on the other hand, we failed to obtain any validity at all. An exception was the unexpected finding that previous training in aviation engineering was significantly but negatively related to success as a pilot. However, this seeming paradox can be explained in terms of the fact that most of our pilot candidates are volunteers, whereas the aviation engineers are drafted into the Air Force and many of them may not really want to fly.

On the general issue of methodology in the study of attitudes and motivation, I should like to point out that, up to the present, applied research on these problems has relied almost exclusively on questionnaire methods. There exists, however, a large body of valuable laboratory data on motivation which has been obtained by the classical experimental method of manipulating independent variables. It would seem that applied psychologists working in this area have tended largely to ignore this potentially fruitful approach. Such a methodological preference is difficult to justify, but I should like to point out the possible benefits to be gained from a greater use of classical experimental methods by aviation psychologists concerned with problems of motivation.

Dr. Abraham Carp (U. S. A.):

For several years at the Air Force Personnel and Training Research Center, we have been carrying out studies very similar to those reported by the present authors. One of our findings from longitudinal studies of motivation may be of interest to the group. In general, we have found little reason for optimism as regards the hope that military pilot training will serve to increase flight motivation. Typically, we find—and this is true not only of flying personnel, but of other types as well—that motivation is highest at the time of entry into the training program, and then undergoes a slow but steady decline. This was something of a disappointment to those who had felt that once we got our young American males into airplanes, they would fall in love with flying. Unfortunately, the opposite turns out to be the case. The more experience a pilot has, the less enthusiasm he seems to have for his work.

A problem we have encountered continually in our attitude studies is that of deliberate distortion. This tends to complicate efforts to use attitude measures for purposes of selection. Direct items such as "Do you like flying?" and "Do you like military service?" are obviously very easy to fake. Interestingly enough, however, we have occasionally found very high validities for items of this type—but, only for the "no" responses. The "yes" responses yield considerably lower validities. The apparent reason for this is that, if an applicant for pilot training answers "no" when asked, for example, "Do you plan to make a career of the Air Force?", then that applicant is quite likely to be telling the truth. If he answers "yes," however, he may either be telling the truth or be concealing his true attitudes, because he wishes for other reasons to be accepted for flight training.

A solution might be to develop items of a more subtle or indirect type. I should be interested to know if the authors have done any work on this problem in connection with their attitude studies.

Dr. Louis Guttman (Israel):

I should like to comment on this very interesting problem of the use of attitude measures in prediction. In studies at the Israel Institute of Applied Social Research, we have been finding that standard linear statistical techniques are not always adequate for the treatment of attitude data because the relations of such measures to proficiency criteria often turn out to be nonlinear. For example, we have found in trying to predict success in civilian work that the degree of strength or positiveness with which an attitude is held may bear a curvilinear relation to success. Thus, in some cases, those individuals who hold a given attitude very strongly tend to fail more often than those in the next interval, who are positive, but not, as it were, overly positive in their opinions. The success rate then decreases again for the next interval of endorsement strength, and, finally, becomes lowest at the negative end of the scale.

Another phenomenon we have found it necessary to look out for is the occurrence of interactions among variables. We have found that the magnitude, and even the direction, of certain correlations may be different for different sub-populations of subjects. Thus, for some sub-populations the relations between a certain at-

titude and success may be positive, while for other sub-populations the relation of the same attitude to success may be negative. This means, of course, that the predictive significance of the attitude variable is partly a function of, i.e., is modified by, the standings with respect to other variables of the individuals under study.

Applying this principle to the problems considered in the preceding paper, we might hypothesize, for example, that flight motivation is positively related to flying success among high-aptitude pilot candidates, but negatively related or unrelated to success among low-aptitude candidates who may want very much to be pilots, but who lack the requisite abilities. To test for the existence of such interactions, and to put them to work in the interests of prediction, non-additive multivariate statistical methods must be utilized.

In brief, then, we have been finding some rather complicated relationships in working with attitude data—relationships for which the usual linear assumptions do not hold. It may be, therefore, that the use of inappropriate statistical techniques is partly responsible for the generally low validities which have been obtained in previous research on attitudes.

Reply

Dr. A. de Brisson:*

In connection with the remarks of Dr. Carp, I should like to note that we, too, have observed a decline in the motivation of pilots during the course of flight training. Using the Guttman scales, we found that a considerable deterioration had taken place by the second or third month of school. This phenomenon is probably not specific to aviation, and we shall doubtless find it occurring in many other fields as well.

Several speakers have referred to the use of attitude scales for selection purposes. I should like to point out that this possibility was not mentioned in our paper. It is the biographical inventory keys which were found to contribute to prediction, and are therefore included in our selection battery. Although attitude measures may, indeed, possess a predictive interest, it is their descriptive aspects with which we are essentially concerned: we plan to use these scales primarily for studying attitude changes.

It should be emphasized that both the Guttman scaling and the homogeneous keying methods are extremely onerous to apply. Moreover, they tend to eliminate a great deal of test data in the interests of internal consistency. The predictive contribution of such measures tends to be somewhat less than that of more global instruments. The addition of biographical inventory keys to a selection battery nevertheless permits one to pass from a multiple validity of, say, .45 to one of .50—certainly, a worthwhile gain. This improvement is obtained, however, only at the expense of a great deal of computational labor.

*Translated from the French.

594869 O - 61 - 8

VALIDITY OF THE U. S. AIRCREW CLASSIFICATION BATTERY IN A SAMPLE OF SPANISH PILOTS

José Germain
Department of Experimental Psychology
Superior Council for Scientific Investigation
Madrid, Spain

(Ed. Note: Professor Germain was unfortunately prevented, by the final illness of his mother, from preparing his manuscript in time for circulation to members of the Symposium. The same unhappy circumstances precluded the possibility of his being present at Brussels to introduce his paper personally. However, his able assistant, Dr. José Luis Pinillos, presented an outline of their joint work. Also, the paper has subsequently become available through its publication in the Revista de Psicologia General y Aplicada, of which Dr. Germain is editor: Germain, J. and Pinillos, J. L., Valoración de la "USAF Aircrew Classification Bettery" en una muestra pilotos espanoles. Rev. Psicol. gen. apl., 1958, 13, 551-560. It is chiefly from the latter source, but also from notes taken on the Pinillos presentation at Brussels, that the following brief abstract has been prepared.)

In the summer of 1955, 120 cadets of the Air Militia University were given a battery of ten tests, seven written and three psychomotor. These were administered a few days after the beginning of flight training. The written tests were Spanish adaptations of USAF tests: (a) Dial Reading, CP622A; (b) Table Reading, CP622A; (c) Coordinate Reading (the Spanish form uses circular coordinates); (d) Instrument Comprehension, CI616C; (e) Mechanical Principles, CI903B; (f) Mechanical Information, CI905B; and (g) General Information, CE505F. The psychomotor tests were: (h) the two-hand coordination test (lathe test), (i) rotary pursuit, and (j) visual discrimination reaction time.

Scoring of the tests was accomplished in Madrid and results were kept secret. Scores were transmuted to stanines, and a final pilot stanine was computed, using the 1952 American weights.

Criteria for validation purposes consisted of the ratings assembled by the Chiefs of Flight Training from the two Schools of the Base, each being responsible for obtaining from subordinates ratings on 60 students. Biserial correlations between test scores and flight records turn out to be very much the same for the Spanish sample as those found in American experience. The pilot stanine shows a validity of .52 for this group, thus comparing favorably with the U. S. standards for primary training. Apparently, the Spanish adaptation of the aircrew battery is an effective one.

Abstract

Validité de la batterie de classification d'équipage (Etats-Unis) sur une échantillon de pilotes espagnols

José Germain
Département de Psychologie expérimentale,
Conseil supérieur de la recherche scientifique
Madrid, Espagne

(Note de l'éditeur: Le Professeur Germain fut malheureusement empêché, par la maladie fatale de sa mère, de terminer son manuscript à temps pour qu'il puisse être distribué aux membres du Symposium. Les mêmes circonstances malheureuses ne lui permirent pas de présenter personnellement sa communication à Bruxelles. Cependant, son assistant, le Docteur José Luis Pinillos, exposa les points principaux de leur travail commun. De plus, cette communication a été publiée depuis dans la Revista de Psicologia General y Aplicada, dont le Docteur Germain est éditeur: Germain, J. et Pinillos, J. L., Valoración de la "USAF Aircrew Classification Bettery" en una muestra de pilotos españoles. Rev. Psicol. gen. apl., 1958, 13, 551-560. C'est principalement à partir de cette source, mais également de notes prises pendant l'exposé du Dr. Pinillos à Bruxelles, que le bref résumé suivant a été préparé.)

Pendant l'été 1955, 120 cadets de l'Université de la Milice de l'Air ont subi une batterie de dix tests, dont sept étaient écrits, et trois psychomoteurs. Ces tests furent administrés quelques jours après le début de l'entraînement de vol. Les tests écrits étaient des adaptations espagnoles de tests de l'U. S. Air Force: (a) Lecture de cadrans, CP622A; (b) Lecture de tables, CP622A; (c) Lecture de coordonnées (la version espagnole utilise des coordonnées circulaires); (d) Compréhension des instruments, CI616C; (e) Principes de mécanique, CI903B; (f) Information mécanique, CI905B; et (g) Information générale, CE505F. Les tests psychomoteurs étaient: (h) le test de coordination bimanuel (test du tour), (i) la "poursuite circulaire," et (j) le temps de réaction à la discrimination visuelle.

La cotation des tests fut faite à Madrid, et les résultats furent gardés confidentiels. Les cotes furent transformées en stanines, et un stanine final de pilote fut calculé en utilisant les poids américains de 1952.

Les critères de validation consistèrent en les appréciations rassemblées par les chefs d'entraînement de vol des deux écoles de la base d'aviation, chacun d'eux étant responsables de l'obtention de ses subordonnés d'appréciations sur 60 élèves. Les corrélations bisérielles entre les résultats de tests et les performances de vol s'avèrent voisines, pour l'échantillon espagnol, de celles trouvées en Amérique. Le stanine de pilote montre une validité de 0,52 pour ce groupe, chiffre qui se compare favorablement avec les standards des Etats-Unis pour l'entraînement primaire. Il apparaît que l'adaptation espagnole de la batterie d'équipage est effective.

General Discussion

Major J. Termöhlen (Denmark):

There are one or two questions I should like to ask about the findings and procedures of this investigation. It is reported in the paper that 120 men entered the flight-training program, and 108 completed the course three months later. I was rather surprised to find a failure rate of only 10 per cent. In Denmark, we commonly lose about 50 per cent of our pilot candidates during the course of training, and we accordingly give a good deal of attention to the reasons for such failures. In fact, in 1956 we instituted the policy of interviewing every pilot who failed to complete training, in order to obtain additional information on causative factors. I should, therefore, like to ask the author if he could tell us why the failure rate for the Spanish sample was so low. Any information on the reasons for the failures which did occur would also be welcome.

The comparison between the Spanish and American validities was also of great interest to me. In view of the differences in populations, training methods, types of equipment, and so on, it was most interesting to observe such a high degree of similarity between the results obtained in the two countries.

Mr. V. Coucheron-Jarl (Norway):

Since we have had a report on the validity of the U. S. Aircrew Battery in southern Europe, I thought it might be of interest to describe how the coefficients came out further north. It is rather remarkable, but in Norway we obtained multiple correlations of up to .53—a value virtually identical to that obtained for the Spanish sample. Our multiple R, however, was based on a total of only six tests. All were of the paper-and-pencil variety, none of the psychomotor tests being included. The sample consisted of 184 pilot candidates, of whom 95 failed to pass the training.

Reply

Dr. J. L. Pinillos (Spain):

Dr. Germain, the author of the preceding paper, unfortunately could not come to the Symposium, so it is I who must attempt to reply to the discussants.

The question raised by Major Termöhlen regarding the number of failures is easily answered. The failure rate was not 10 per cent; it was around 40 per cent of the total number of individuals rated. The figure given in the paper refers to the number of candidates who voluntarily dropped out during the course of training, and who therefore were not rated by the chiefs of the schools. Of the 108 candidates who remained in training, approximately 40 per cent did not pass the requirements. So, it can be seen that the failure rate in our respective countries is after all highly similar.

It can be noted parenthetically that since the introduction of the selection battery the number of failures has been declining, and is now down to around 20 per cent.

As to the matter of the high degree of similarity between the Norwegian, American, and Spanish results, I can only say that neither Coucheron-Jarl nor I am responsible for the fact that the selection battery is universally good. We have carried out two other validation studies, not described in the paper, which also yielded correlations similar to those reported. One investigation was based on a sample of 24 candidates, and the other on a sample of 66. Each study yielded criterion correlations, both for the individual tests and for the battery as a whole, which did not markedly differ from the present ones. The multiple validities were around .45, and the individual test correlations in no case differed by more than 10 to 20 points.

EVENING ADDRESS

Dr. Arthur W. Melton, Director

U. S. Air Force Personnel and Training Research Center

Lackland Air Force Base, Texas

U. S. A.

Introductory Remarks

Dr. Frank A. Geldard (U. S. A.):

The speaker of the evening has figured so prominently in military psychology and international scientific affairs that for most of you he probably needs no introduction. In 1941, when the United States entered World War II, Dr. Melton was brought in from civilian life to participate in the Aviation Psychology Program of what was then the U. S. Army Air Corps. He has served military psychology continuously since that time. When others were beating their swords back into plowshares after the war, Dr. Melton, then Professor at Ohio State University, was working on what is now the familiar blue volume on psychomotor tests. In 1947 he became chairman of the Panel of Psychophysiology of the Research and Development Board of the U. S. Department of Defense, an important post in which he did much to aid the post-war development of military psychology. In 1949 he once more became actively engaged in work for what by then had become the United States Air Force. At present, as all of you know, he is director of the Air Force Personnel and Training Research Center at Lackland Air Force Base, Texas. With this heavy responsibility, he nevertheless manages also to serve as editor of one of our liveliest journals, The Journal of Experimental Psychology. For about 16 years, then, Dr. Melton has been actively and continuously serving the cause of military psychology. I think it is safe to say that throughout this period he has contributed more conspicuously to the field than anyone else in America. It is with the greatest pleasure that I introduce him tonight to give us an account of the present status of military psychology in the United States.

Military Psychology in the United States of America

Arthur W. Melton
Air Force Personnel and Training Research Center
Lackland Air Force Base, Texas
U. S. A.

Mr. Chairman, ladies and gentlemen, I am greatly honored, and therefore deeply pleased, by this opportunity to present for your consideration an overview of military psychology in the United States. Certainly a heavy responsibility is shouldered by anyone chosen to represent the very large number of scientists and the many laboratories and programs involved in military psychology in the United States today. To describe this effort in a communication short of book length requires, of course, emphasis on generalities, and a great deal of selection among generalities as well. If the generalizations turn out in some instances to be con-

troversial, I am sure that my numerous colleagues at this Symposium can provide the necessary antidotes.

Before proceeding further, it might be well to discuss the meaning of the term "military psychology." In the American Psychological Association, the Division of Military Psychology, which was established in 1944, is made up of psychologists who have worked for the military establishment either as part of the in-house effort or on contract programs. One finds among them psychophysiologists, experimental psychologists, psychometricians, social psychologists, clinicians, and not a few psychologists whose duties are primarily administrative. Today, military psychology is virtually coextensive with all psychology, and is distinguished only by its characteristic contexts of application—such as the operation or maintenance of the fire-control system of an interceptor aircraft, the radar interpretation tasks of a navigator-bombardier, the specifics of a personnel management system that has its roots in the traditions of a service, the unique characteristics of an authoritarian society, or the hazards of combat. The appropriate analogue of military psychology is industrial psychology, which is also defined by specific contexts of application—contexts determined in this case by the needs of the business and industrial worlds.

Two important conclusions can be drawn from these observations. First, military psychology will be heavily dependent for its advancement and effectiveness on the level of psychology in general, and must, in its own self-interest, support certain kinds of fundamental research. It cannot be emphasized too strongly that by the very nature of its formal relations to psychology at large, military psychology depends on progress in basic scientific psychology.

The second conclusion is that if psychologists are to be effective in applying psychology to military problems, they must become familiar with the specific policies, procedures, operations, and weapons of the military service to which the applications are to be made. Insofar as these policies, weapons, and operations differ among the Army, Navy, and Air Force, a distinction must be made between Army psychologists, Navy psychologists, and Air Force psychologists. However, I strongly suspect that the role of the military psychologist, as well as the requirements for and applications of research and development in military psychology, are essentially the same in all three services. At least, I shall make this assumption when I develop, somewhat later, the place of the concepts of the personnel management system and the weapon system in recent developments in military psychology in the United States.

One final statement about military psychology, as viewed here, is perhaps in order. It should be emphasized that the present concern is with military psychology as a branch of psychological science and technology, not as a professional occupation of psychologists in or out of uniform. Only if the activity of the psychologist in the military establishment involves research, development, testing, and inferential extrapolation from systematic knowledge and theory do I intend to include it in what I am discussing.

With these comments as background, let me now turn to the principle trends and emphases of contemporary military psychology in the United States. These trends, of course, have their roots in the past, going at least as far back as World War I. American psychologists of that day not only produced the Army Alpha and Army Beta General Classification Tests, they also involved themselves in the development of job-knowledge tests, the training of naval gunners, the analysis of

aircraft pilot ability, and probably in many other familiar-sounding military problems about which there is no written record.

Between World Wars I and II, there was almost no interest shown by American psychologists in military problems, perhaps because there was almost no interest on the part of the military in obtaining their assistance. Exceptions were the development of a new and improved Army General Classification Test by the Army Adjutant General, and some research on psychomotor selection tests for aircraft pilots carried out under the aegis of the Surgeon of the Army Air Corps. However, when the clouds of World War II formed on the horizon, psychologists were recruited rapidly and in large numbers to conduct a great variety of research studies and supervise technical applications in military operations.

The history of this effort is so recent, and so familiar to most of you, that little needs to be said about it. We might, however, comment on the way in which psychologists were put to work by the military. First, it can be noted that the largest single program, that of the Army Air Forces, was organized under medical auspices, and emphasized problems involving the psychological selection of personnel almost to the exclusion, at least until late in the war, of training problems, proficiency measurement, and many other important aspects of the military personnel system. Other programs, such as those under certain Bureaus of the Navy, the Adjutant General of the Army, and the National Defense Research Council, were not thus restricted. Second, it is my impression that military requisitioning of psychological assistance tended to be on a piece meal basis—on what we in the States call a "fire-fighting," a "snake-killing," or, more elegantly, a "crash" basis. This is to be contrasted with systematic utilization of psychological science and technology, an orientation which has become increasingly characteristic of military psychology in the United States during the past ten years, and which forms the keynote of my discussion tonight.

However, before attempting to define this unfolding systematic role of psychologists in military affairs, I should like to convey to you something of the magnitude of our effort in military psychology during the last decade. My technique was a hasty counting operation, undoubtedly inaccurate, but good enough for the present purpose. In the 1948 Directory of the American Psychological Association, I identified 98 psychologists who were working for the Departments of Defense, Army, Air Force, or Navy. This represents about 2 percent of the 5,047 psychologists who constituted the membership of the Association at that time. In the 1957 APA Directory, by contrast, I counted 729 psychologists working for agencies of the military departments. (This number, even though it includes those employed by the RAND Corporation, a large contract agency of the Air Force, and by the Human Resources Research Office, a large contract agency of the Army, is probably an underestimate as compared with the 1948 figure, since I counted only individuals working for agencies that employed three or more psychologists.) These 729 psychologists constitute almost 5 per cent of the fifteen thousand members listed in the 1957 Directory.

The relative extent of this effort can be better appreciated when it is known that, again according to hasty count, approximately 33 per cent of the psychologists listed in the 1957 Directory hold positions in universities or colleges. Currently, then, there is one American psychologist in full-time government service (military or civilian) for every 6.5 academic psychologists. Another clue to the magnitude of the effort, perhaps of special interest to my compatriots, is the fact that American military psychologists, by my counting system, now outnumber clinical psy-

chologists in the government Veterans Administration program by a factor of seven to six. Even this information has two cutting edges, since many of the clinical psychologists in the Veterans Administration are concerned with fundamental research on problems of interest to military psychology, and, perhaps, properly conceived as a part of military psychology.

The in-house effort in military psychology is, of course, only a part of the total effort. In the first place, the work of these seven-hundred-odd psychologists is supported by many non-psychologist officers and men of the armed services, and by civilian engineers, mathematicians, technicians, and clerks. In the second place, there exist contract research-and-development programs in military psychology and related areas which have had a total budget of approximately five million dollars for the last several years. A very great part of this contract fund goes to colleges and universities, and much of it is for the support of fundamental research on problems of importance to the military.

Parenthetically, it can be noted that these data on the in-house and contract efforts of our military establishment strongly suggest that the in-house effort outweighs the contract effort. If one assumes that each in-house psychologist represents an expenditure of $25,000 for salary and supporting personnel and equipment (which is probably an underestimate), then the in-service programs involve an expenditure of some eighteen million dollars per year. This is to be compared with the five million per year in contract funds. Even though there are undoubtedly some errors in my accounting, the balance would seem to be surely and heavily in favor of the in-house effort. In my opinion, this situation has been highly favorable to the development of military psychology as a self-conscious discipline. It is even possible that a systematic program of application of psychology to military problems must have a strong in-house effort as a major support. However this may be, there can be no questioning of the fact that American military psychology is at present highly institutionalized.

The Army has, among other agencies, the Personnel Research Branch of the Adjutant General's Office, the Human Resources Research Office (a contract effort primarily in support of the Continental Army Command), the Psychology Department of the Army Medical Research Laboratory, the Psychology Department of the Walter Reed Army Institute of Research, the Army Ordnance Human Engineering Laboratory, and psychology sections in the Quartermaster Research and Development Command.

The Navy has an Engineering Psychology Branch in the Naval Research Laboratory, a Human Factors Division in the Naval Electronics Laboratory, a Personnel Analysis Division in the Bureau of Naval Personnel, and psychological units as parts of, among others, the Naval Medical Research Institute, the Aeronautical Instruments Laboratory, the Aircrew Equipment Laboratory, the Aviation Medical Acceleration Laboratory, the Medical Research Laboratory, and the Office of Naval Research. This, although an incomplete list, will give some notion of the wide dispersal of military psychologists in the Navy.

In the Air Force, all military psychological research and development is carried out within the Air Research and Development Command, except for a Department of Psychology and a Department of Clinical Psychology in the Air Force School of Aviation Medicine, and a small element in the Arctic Aero Medical Laboratory. Within the Air Research and Development Command, there are human engineering research and development agencies at Wright Air Development Center, Rome Air

Development Center, the Air Force Cambridge Research Center, the Air Force Special Weapons Center, and the Air Force Flight Test Center. All personnel and training research in the Air Force is currently allocated to the Air Force Personnel and Training Research Center, which is one of the eleven major operating agencies of the Air Research and Development Command.

The first purpose in this lengthy listing of psychological research installations in the military establishment was to give appropriate emphasis to the great diversity of uses to which psychologists are put, and to the great variety of the projects for which, as is apparent from their very titles and associations, the various agencies are responsible.

The second purpose was to indicate roughly how military psychology is organized within the U. S. Armed Forces. In the Navy, small groups of psychologists are located at many different places, with their responsibilities defined by some parent organization. There thus exists here a rather strong decentralization of mission and responsibility. The Army, although its program also tends toward decentralization, has endowed two agencies (the Psychology Branch of the Adjutant General's Office and the Human Resources Research Office) with rather broad responsibilities in military psychology. Finally, in the Air Force one finds the control of military psychology firmly centralized in the Air Research and Development Command, and, within that Command, all personnel and training research, i.e., virtually all military psychology except "human engineering," is concentrated in one organization: the Air Force Personnel and Training Research Center.

I do not propose to argue the relative merits of centralized versus decentralized organization in the management of military psychology, but I do believe that a centralized organization, such as the Air Force has, definitely encourages a systematic conceptualization of the role of psychology in military affairs. And I further believe that this systematic conceptualization of the role of military psychology is a ripple that will in the future become a wave.

The needs of the military establishment for psychological knowledge can be grouped under three headings: (a) improvement of the military personnel management system, (b) contributions to the development and effective operational use of new weapons and weapon systems, and (c) analysis of the interactions of the personnel management and weapon systems.

My thesis is that these unifying concepts of "weapon system" and "personnel management system" can be used not only to organize and assess the past contributions of military psychology, but also to identify the research and development needs of the future. My subordinate thesis is that the use of these organizing concepts to define the role of psychology in military affairs is one of the most significant recent developments in military psychology in the United States.

While the Air Force has been the most vocal, the most explicit, and the most self-conscious of the three services in the use of these concepts, it would be wrong to think it alone in this development. One can find historical precedents in the Navy psychologist's approach to the submarine as a man-machine system, and in his approach to the functions and manning of combat information centers. There is also some evidence for use of the concept of the personnel management system as a frame of reference in Navy research on personnel problems. The Army seems occasionally to employ the personnel management system concept, but uses of the weapon system concept in Army psychological work have been difficult for me to

locate. This may be due to the fact that the Army has only recently begun to deal with man-machine systems of a complexity approaching that typical of Navy and Air Force systems.

I should like to discuss first the concept of the military personnel management system—that entire complex of policies, procedures, and operations involved in supplying the military forces with human beings possessing the characteristics necessary to effective performance in combat and combat-support operations. The raison d'etre of the military personnel system exists in the need to select capable men, and to train them in the roles required by weapons which are becoming increasingly more complex integrations of men and equipment. The military personnel system can be said to have the function of producing the human components of each weapon system to specifications that are inherent in, or directly derivable from, the design and intended operational use of the weapon system.

We find this role of the personnel system gaining greater recognition within the military services as weapons become more complex, as the demands upon human skills, knowledge, and decision-making capacities become heavier, and as the failures of traditional military personnel practices become accordingly more devastating to combat effectiveness. Evidence of this recognition is the fact that the Air Force now officially defines a weapon system—such as an interceptor aircraft and its associated ground operations—as equipment plus the knowledge and skills required to operate and maintain it. One therefore finds the personnel specialist in the Air Force becoming increasingly concerned with the specific characteristics of new weapons under development.

What are the elements of a personnel management system that should become focal points in the research and development programs of military psychologists? Some of the major elements are the following: (a) military organization, (b) personnel management, (c) occupational analysis (including job and task definitions and descriptions), (d) national and military manpower analysis, (e) personnel procurement, (f) personnel classification, (g) formal training, (h) personnel assignment and reassignment, (i) on-the-job training, (j) instructional devices, (k) handbooks, job aids, and standing operating procedures, (l) personnel proficiency measurement, (m) cross-training and career development, (n) maintenance of personnel effectiveness under unusual social and environmental conditions, (o) retention and re-enlistment of personnel, and (p) retirement of personnel.

This list is certainly without novelty. For many years, such elements have been used by military psychologists to describe the military "personnel operations" to which their research applies. The difference in our present approach lies in its emphasis on the interrelations of the elements, and, hence, on the necessity, when necessity, when conducting research and recommending applications, of giving full recognition to their interactions. For example, in the past few years it has become increasingly evident that one cannot define military jobs without making certain critical assumptions about military organization. Similarly, one cannot approach training research without giving immediate attention to the potential interactions between training methods and personnel selection and/or proficiency standards. Again, there is now recognition of the fact (after the fact!) that research on the design of handbooks and other job aids must reflect the intimate interactions between the characteristics of such handbooks or aids and, for example, the quality of the recruited personnel, the way in which the military jobs are structured, and the level of training and education given to the individuals who will use the handbook or aid. Conceptualization of military personnel management as

a system of operations designed to select and train men to match the machinery of war would thus appear to constitute a more efficient and effective way of structuring the role of behavioral science in this endeavor.

The value of the concept of the personnel system becomes even clearer when one attempts to formulate long-range programs of research in military psychology. The time lags involved mean that one must anticipate by many years the critical personnel problems of the future, match them against anticipated developments in science and technology, and identify those areas of basic long-range research that the military establishment must support in order to insure solutions of its future problems. Some of us have found that attempting to do this sort of thing is virtually impossible without the personnel management system as a conceptual framework. In fact, we have even found it necessary to go beyond defining such a system solely in terms of military goals, and to consider interactions between the military personnel system and the national social system in which the military organization is embedded. But, extension of my discussion along these lines would lead me too far afield, so I shall return to my main theme by pointing out that clues as to the nature of the military personnel system in, say, 1965, can best be obtained from an analysis of the anticipated operating characteristics of weapon systems considered as possibilities for 1965. Accordingly, I should now like to discuss the role of the military psychologist in the development of new weapon systems.

The psychologist becomes involved in the development of a new weapon system in two important ways which should not, in practice, be separated. First, the psychologist is the scientist primarily responsible for determining the changes which will be necessitated by the new weapon system in the supporting military personnel system. Second, he is and must be intimately involved in the design of the weapon system, since he is the scientist who knows most about the capacities and limitarions of the human components which must be married to the equipment components to produce the new system.

It will readily be deduced that the first responsibility encompasses those mentioned in my preceding discussion of the military personnel system, and that the resulting research problems are similar to the ones just described. There is, however, a difference. In the present case, the psychologist is in the position of having to specify the "personnel subsystem" required to produce satisfactory human components for the new weapon system far in advance of the creation of even a first model of that system. The reason for this is that the personnel requirements must be fulfilled according to a schedule which will bring the human components together with the hardware components as soon as the weapon system is ready for tests and operational use.

Only through bitter experience with delays sometimes exceeding two full years, have we come to realize that new weapon systems almost invariably generate new jobs or critical revisions of old jobs, and that this necessitates revision of the existing personnel system. New selection criteria, new job classifications, new training courses and tailor-made training devices, new on-the-job aids, new proficiency measures—any or all of these may be needed. Thus, the psychologist must collate general psychological principles with information about equipment design, the intended operational use of the weapon system, and the existing characteristics of the military personnel system to produce an appropriate personnel subsystem before the new weapon system comes into existence.

This function of psychologists in the military establishment has three important implications for military psychology as a science. First, the requirement for

extrapolating psychological knowledge to new and untried systems underscores the need for generalizable research on the previously discussed elements of the personnel system. Second, the anticipated characteristics of future weapon systems serve as a source of information on which to base decisions as to the relative priority of needs for research in military psychology. Finally, the role of the military psychologist in developing personnel systems for new weapons has established a firm basis for man-machine systems analysis and research as an area of military psychology.

Let us turn now to a consideration of the psychologist's two roles in weapon system development. It is anticipated that although future weapon systems will involve greater automation of operations and maintenance, the functions of the human component will nevertheless remain critical. If the psychologist is to provide effective guidance for the allocation of systems operations to men and machines, it may prove necessary to abandon the distinction between "human engineering" (which has been concerned with the effects of human capabilities and limitations on the design of equipment components) and "personnel and training research" (which has been concerned with the selection, training, and quality control of the human operator). The effective integration of human components into weapon systems clearly requires a unified approach. Analysis of the interaction of human and equipment components, of the availability and trainability of human operators, and of their maintenance at par as elements of the system, must be carried out simultaneously at all stages of weapon system design, development, and use.

To date, almost all research on the interactions of human and equipment components has been concerned with abstracted portions of relatively simple systems, such as the pilot and fire-control subsystems in an interceptor aircraft. This is because an adequate theory and methodology for the global analysis of complex systems does not yet exist. As a consequence, it is not presently possible to specify quantitatively the optimal arrangement of human and hardware components for a complete system, nor to identify the overall changes in system effectiveness resulting from improvement or deterioration in specific human or equipment components. It is evident that techniques for total systems analysis are required for effective "human engineering," and for exact specification of the human knowledge and skills that must be produced through selection and training.

It is believed that the military establishment must initiate and support research on the theory and techniques of systems analysis immediately if future weapons systems are to reflect adequate consideration of the human components of systems. In view of recent contributions to this effort, I estimate that it may not be too many years before we have answers to the following questions: (a) What is the optimal division of system functions between human and equipment components? (b) How must the hardware components be designed if the system is to provide for the reinforcement (reward, maintenance of motivation, and skill) of the human component? (c) When the system involves multiple human and multiple hardware components, what are the optimal linkages between man and man, and between man and machine? (d) What are the essential conditions for systems training, and the required characteristics of system training devices? (e) How can the operational readiness of systems be defined and verified in the absence of actual wartime operations?

Both engineering and psychology are ready for a frontal attack on these problems if the necessary facilities and resources are provided. It is encouraging that several efforts in this direction are already underway in the United States.

Finally, I need but note again that each new weapon system brings with it certain requirements for change in the overall military personnel system. It should therefore be clear that the military personnel management system of some specific future date is to a substantial degree predictable if one knows the characteristics of the weapon systems under development today, and proposed for use at that date. Thus, the psychologist must participate in long-range planning for optimal military personnel systems of the futire, and for the evolutionary and sometimes revolutionary changes that may have to be made in order to achieve them.

My conclusion from all this is simple: I believe that military psychology is here to stay—even though it may have its unpopular and unfunded moments. Psychology is too useful—even too necessary—to be ignored by the military establishment.

Summary

Military psychology is today virtually coextensive with psychology in general, being distinguished only by the characteristic nature of the problems to which it is applied. It is concluded that: (a) military psychology depends for its advancement on progress in basic scientific psychology, and (b) psychologists must become familiar with specific military policies, procedures, operations, and weapons if they are to be effective in dealing with military problems.

A brief historical review of the development of military psychology in the United States is then presented. During World War I, American psychologists produced the Army Alpha and Beta General Classification Tests and carried out research in the areas of training, job analysis, and the development of job-knowledge tests. Between wars their interest in military problems tended to lapse, but was abruptly revived at the beginning of World War II, when psychologists were recruited rapidly and in large numbers to engage in extensive programs of basic and applied research. During the war, military utilization of psychological science and technology tended to be on a piece-meal basis—an orientation which is to be contrasted with the more systematic approach of the last ten years. Today, American psychologists are widely and integrally involved in military affairs.

An idea of the magnitude of the contemporary American effort in military psychology can be gained from the fact that the number of psychologists working for military agencies has risen over the last decade from approximately 2 per cent to approximately 5 per cent of the total membership of the American Psychological Association. There is approximately one American psychologist employed in full-time military or civilian government service for every 6.5 academic psychologists, and military psychologists now outnumber clinical psychologists in the federal Veterans Administration program. In addition, this extensive in-house effort is supplemented by a large government program of contract research in military psychology and related areas.

It is estimated that the in-house effort outweighs the contract effort by approximately 3.5 to 1. American military psychology can accordingly be said to be

highly institutionalized. The various research installations maintained by the Army, Navy, and Air Force are listed as a means of indicating: (a) the scope and nature of the research areas currently being supported by the military, and (b) the way in which military psychology in the United States is organized. Centralization of control, which is strongest in the Air Force, is seen by the speaker as having contributed to the current systematic conceptualization of the role of the psychologist in American military affairs.

In discussing the needs of the military establishment for psychological research, the organizing concepts of the personnel management system and the weapon system are introduced. It is contended that the function of military psychology is to contribute to the improvement of both systems and to understanding of their interactions with each other. The recognition and use of these unifying concepts is regarded as one of the most significant recent developments in American military psychology.

The purpose of the military personnel management system is to produce, by appropriate selection and training, human operators possessing the skills and attributes necessary for performing effectively as components of complex weapon systems. By thus conceptualizing the various areas of personnel research in terms of a system of operations designed to match men to the machinery of war, attention is focused on: (a) the relations of the various areas to each other and to weapon systems, and (b) the necessity of taking such inter-relations into account in planning and carrying out research.

Turning to a consideration of weapon systems, it is pointed out that the military psychologist must contribute in two ways to the development of new weapons. First, he must determine what changes in the supporting personnel system will be necessitated by a new weapon system, and second, he must participate in the designing of the weapon system itself.

The first responsibility is complicated by the fact that if costly delays are to be avoided, the psychologist must determine the personnel requirements of new weapons far in advance of their creation. This can best be done by analyzing the operating characteristics of weapon systems planned for the future, and attempting to deduce the personnel operations which will be required to produce appropriately trained human components.

The second responsibility involves insuring that psychological knowledge of human capacities and limitations is taken into account in the design of weapon systems which depend, for their effectiveness in part, on the activities of human operators. It is suggested in this connection that since efficient integration of human components into a weapon system is a problem requiring a unified conceptual approach, it may prove desirable to abandon the currently accepted distinction between "human engineering" and "personnel and training research." The intimate interactions between the human and equipment components of a weapon system require that attention be given simultaneously to problems of personnel selection and training, and human-engineering research on the implications of human capacities for equipment design.

It is pointed out that, because an adequate theory and methodology for global analysis of complex systems does not yet exist, research on the interactions between human and equipment components has to date been concerned exclusively with abstracted portions of relatively simple systems. Techniques for total systems

analysis are sorely needed if human-engineering and personnel research is to yield optimal results. Hence, the military establishment must initiate and support basic research on the theory and techniques of systems analysis. Problems of special importance in this field are those of: (a) specifying the optimal division of system functions between human and equipment components, (b) determining how equipment can be designed to reinforce the human operator and thus maintain his performance at maximum efficiency, (c) determining the optimal linkage between the human components and between the human and equipment components of a system, (d) identifying the essential conditions for systems training and the required characteristics of system training devices, and (e) defining and verifying the operational readiness of a system in the absence of wartime conditions.

The current role of the psychologist in the military establishment has three important implications for military psychology as a science. First, the fact that existing knowledge must be extrapolated to new systems underscores the need for generalizable research on personnel problems. Second, decisions as to the relative priority of basic research needs can be made in part in terms of the anticipated characteristics of future weapon systems. Third, the contribution of the psychologist to problems of system design and development has established this research field as an integral area of military psychology.

In conclusion, it is emphasized that since each new weapon system requires changes in the supporting personnel system, the psychologist must participate in long-range planning in the domain of personnel and systems research.

Résumé

Psychologie militaire est aujourd'hui virtuellement coextensive avec la psychologie en général, ne s'en distinguant que par la nature caractéristique des problemes auxquels elle s'applique. Il faut en conclure que (a) la psychologie militaire dépend, pour son propre avancement de celui de la psychologie scientifique fondamentale, et que (b) les psychologues doivent se familiariser avec les principes, les méthodes, les opérations et les armements militaires s'ils tiennent à faire un travail efficace lorsqu'ils s'occuperont de problèmes militaires.

Un bref historique de l'évolution de la psychologie militaire aux Etats-Unis est alors présenté. Au cours de la première guerre mondiale, les psychologues américains ont développé les tests de classification générale "Alpha" et "Beta" de l'Armée, et ont mené des recherches sur l'entraînement, l'analyse du travail et la mise au point de tests de connaissance du métier. Entre les deux guerres, leur intérêt pour les problèmes militaires a montré une tendance à s'affaiblir, mais a été brutalement ravivé au début de la deuxième guerre mondiale, quand des psychologues ont été recrutés en hâte et en grand nombre pour s'atteler à des programmes poussés de recherche théorique et appliquée. Pendant la guerre, l'utilisation militaire de la science et la technologie psychologique tendaient à s'improviser au hasard de chaque cas, attitude contrastant avec l'orientation plus systématique des dix dernières années. Aujourd'hui, les psychologues américains sont largement, et intégralement, engagés dans les affaires militaires.

On se fera une idée de l'ampleur de l'effort américain contemporain dans le domaine de la psychologie militaire en considérant que le nombre des psychologues travaillant pour des institutions militaires a crû au cours des dix dernières années d'environ 2 pour cent à environ 5 pour cent de l'effectif total de l'American Psychological Association. On compte approximativement un psychologue américain employé par le Gouvernement à titre militaire ou civil pour 6,5 psychologues académiques, et le nombre des psychologues militaires dans le cadre du programme fédéral de la Veterans' Administration dépasse à présent celui des psychologues cliniciens. En outre, cet effort interne est complété par un vaste programme gouvernemental de contrats de recherche en psychologie militaire et d'autres domaines apparentés.

On estime cet effort interne supérieur à l'effort contractuel dans une proportion d'environ 3,5 à 1. La psychologie militaire américaine peut donc se dire fortement institutionnalisée. Les diverses installations de recherche de l'Armée, de la Marine et de la Force Aérienne sont cataloguées pour fournir une indication quant à: (a) l'étendue et la nature des domaines de recherche actuellement appuyés par le secteur militaire de (b) la façon dont la psychologie militaire est organisée aux Etats-Unis. La centralisation des contrôles, que la Force Aérienne pousse le plus loin, est vue par l'orateur comme facteur contributant à la notion systématique du rôle actuel du psychologue dans les affaires militaires américaines.

Une discussion des besoins de l'institution militaire en matière de recherche psychologique introduit les notions directrices de système de maniement du personnel et de système d'armements. La psychologie militaire aurait pour rôle, prétend-on, de contribuer à améliorer les deux systèmes et à saisir leur interaction mutuelle. La reconnaissance et l'utilisation de ces deux notions unificatrices est considérée comme l'un des développements récents les plus significatifs dans le domaine de la psychologie militaire américaine.

Le système militaire de personnel a pour objet de fournir, par une sélection et un entraînement appropriés, des opérateurs humains dotés des attributs nécessaires à l'exécution efficace de leurs fonctions comme éléments composants des systèmes complexes d'armements. En ramenant ainsi les divers domaines de la recherche en matière de personnel à la notion d'un système d'opérations conçues pour adapter les hommes au matériel de guerre, l'attention est attirée sur: (a) les rapports des divers domaines entre eux et avec les systèmes d'armements, et (b) la nécessité de tenir compte de ces rapports dans la programmation et le déroulement des recherches.

Passant alors à l'examen des systèmes d'armements, il est fait observer que le psychologue militaire doit contribuer de deux façons à la mise au point de nouveaux armements. Tout d'abord, il doit déterminer les modifications qu'un nouveau système d'armements nécessitera dans le système du personnel de soutien; en second lieu, il doit participer à la conception du nouvel armement même.

La première de ces deux tâches se complique du fait que, pour éviter des retards coûteux, le psychologue doit déterminer les besoins en personnel pour les nouveaux armements longtemps avant l'introduction de ceux-ci. La meilleure formule à cet égard consiste à analyser les caractéristiques opératoires des systèmes d'armements prévus pour l'avenir, et à s'efforcer d'en déduire les procédés en matière de personnel exigé pour fournir des éléments composants humains convenablement formés.

La seconde tâche consiste à assurer que la connaissance psychologique des pouvoirs et des limitations humains soit prise en considération dans l'élaboration de systèmes d'armements dont l'efficacité dépend en partie du comportement des opérateurs humains. A cet égard, il est avancé que, une intégration efficace des éléments humains aux systèmes d'armements exigeant une approche conceptuelle unifiée, il peut être souhaitable de renoncer à la distinction couramment admise entre "human engineering" ("génie humain") et la recherche sur le personnel et la formation. Les interactions intemes entre les composantes humaines et les composantes matérielles d'un système d'armements imposent le souci simultané des problèmes de sélection et de formation du personnel, et des recherches de "génie humain."

Il est fait observer que l'absence d'une théorie et d'une méthodologie convenables à l'analyse globale des systèmes complexes a jusqu'ici limité les recherches sur les interactions entre composantes humaines et composantes matérielles à certaines parties de systèmes relativement simples. Il faut élaborer des méthodes d'analyse totale si l'on veut que les recherches en matière de "génie humain" et de personnel comportent des résultats optimaux. C'est pourquoi l'institution militaire doit entamer et soutenir la recherche fondamentale sur la théorie et la méthodologie de l'analyse des systèmes. Parmi les problèmes particulièrement importants à cet égard on peut citer: (a) la spécification de la répartition optimale des fonctions des systèmes entre composantes humaines et composantes matérielles, (b) la façon de concevoir l'équipage pour qu'il soutienne l'opérateur humain et maintienne par là son rendement à un maximum, (c) la détermination du couplage optimal des composantes humaines entre elles et avec les composantes matérielles d'un système donné, (d) l'identification des conditions essentielles et des caractéristiques nécessaires des dispositifs de l'entraînement et (e) la définition et la vérification de l'applicabilité opérationnelle d'un système en l'absence de l'état de guerre.

Le rôle du psychologue dans le cadre de l'institution militaire met en relief trois aspects de la psychologie militaire en tant que science. En premier lieu, la nécessité d'extrapoler les connaissances existantes à des systèmes souligne l'importance de recherches généralisables sur les problèmes de personnel. En second lieu, les décisions quant à la priorité relative des besoins en recherches fondamentales doivent être prises partiellement sur la base des caractéristiques théoriquement prévues des futurs systèmes d'armements. En troisième lieu, la contribution du psychologue à la résolution des problèmes de conception et de mise au point des systèmes a élevé de domaine de la recherche au rang de partie intégrante de la psychologie militaire.

En conclusion, on soulignera que, tout nouveau système d'armements commandant des modifications du système du personnel de soutien, le psychologue doit participer à une programmation à longue échéance des recherches en matière de personnel et de systèmes.

SESSION III

Chairman: Dr. R. W. van de Giessen

Royal Dutch Navy

The Hague, The Netherlands

SELECTION OF ARMY OFFICERS FOR COMBAT SERVICE

Leonard Carmichael
Smithsonian Institution
and
Lynn E. Baker
U. S. Army Research and Development Office
Washington, D. C.

The contributors of psychology in general, and psychometrics in particular, to the selection of men to be trained as United States Army officers for the defense needs of the future have been considered in recent months by the authors and some of their colleagues. Central attention has been given to the novel demands that will be made on commissioned officers under combat conditions by the new tactics emerging as a result of the use of modern weapons.

As a basis of the study, the fact was kept in view that the range of functional capacities of human sense organs, brains, and muscles has probably not varied during historic times. The pattern of qualities required of effective military leaders, in the past, such as outstanding muscular power, has altered with the various major changes in weapons and the resulting modification in tactics and manpower requirements. Of particular interest to the psychologist at the present time are the greatly increased requirements of technical knowledge to assure the maintenance and use of the weapons, and the greatly increased demands on the commander's abilities for independent judgment and personal leadership in their tactical exploitation.

In the more leisurely past centuries, the new criteria for the selection of officers with the new capabilities required by new weapons and procedures could have grown by gradual trial and error. Today's emphasis on speed in organizing a large fighting force and the great expense of training officers for it underlines the importance of using all appropriate modern scientific techniques in identifying characteristics needed by officers in modern types of combat. It is necessary also to develop procedures to measure the extent of these required qualities in the civilians to be selected and trained as officers. It has become increasingly clear that the informed judgments of qualified peers and superiors are of great value in determining which candidates can be developed into effective officers. It is equally clear, however, that such selections are better when they are based on categories of judgment whose relevance has been specified and validated, and when they are reinforced by appropriate psychometric findings and other modern assessment procedures. To put it another way, the demands that must be made on junior commissioned officers at the present time and in the immediate future are so much more complex than were requirements only a few years ago that methods effective in the selection of officers for the more conventional armies of the nineteenth century must now be reinforced by the use of modern personnel tools, if complete success is to be expected.

The Armed Forces Qualification Test (AFQT) has been jointly developed through three revisions by all U. S. military services. It has demonstrated itself to be useful in predicting the general "trainability" of men in the military situation, and its administratively determined cut-off score establishes a level below which individuals are unacceptable for military service. Following screening by AFQT, the Army Classification Battery (ACB) gives factorial scores which can be combined to estimate specific "Aptitude Areas" relevant to particular military occupations. The seven "Aptitude Areas," or patterns of aptitudes, currently used by the Army for classification of individuals for enlisted men's specialties are: the Combat Aptitude Area, a General Technical Aptitude Area, a Motor Maintenance Aptitude Area, a General Maintenance Aptitude Area, a Clerical Aptitude Area, a Radio Code Aptitude Area, and an Electronics Aptitude Area. Utilization of these Aptitude Areas in assigning men to advanced training for military specialties has the practical effect that a man whose AFQT score is, say, 100 performs in the area of his best aptitudes with a proficiency which otherwise would require an AFQT score of 119.

When facing the problem of selecting men with the qualities that will allow them to be trained rapidly and effectively as junior officers in Army combat forces of the future, it is all very well to say abstractly, if the performance can be measured the capacity for future performance can be predicted. But, in this complex performance exactly what are the characteristics of young civilians that should be measured? In the U. S. Army, the Adjutant General's Personnel Research Branch (PRB) and the Human Resources Research Office (HumRRO) have, it is true, already identified a number of measurable characteristics which are indicators of the probable value of enlisted men in combat. These measures indirectly help in selecting individuals who have what may be termed modern fighter potential. There is good reason to assume that some of these measures may also help in selecting commissioned officer candidates. Some of the best of these techniques identify and measure characteristics which have been isolated on the basis of reports, observation, and measurement of 4,000 men who were in actual combat in Korea. Further work is being done on these important instruments of selection.

A good current device related to the prediction of combat success of commissioned officers is the Aptitude for Service Rating (ASR) of the final year of the United States Military Academy at West Point. The ASR has been demonstrated to have value when applied at Officer Candicate Schools in identifying the individuals who are likely to fail to complete training.

American military services have analyzed some of the main functions required of junior commissioned officers who must perform very specific functions. On the basis of this analysis, patterns of tests such as the so-called "Pilot Stanine" of the Air Force have been developed which have good reliability and validity. This is also true of other patterns of tests, such as the "Navigator Stanine" of the Air Force.

At the present time it is not possible to report that similar, statistically validated instruments have been developed to select those likely to succeed in training and in later performance as combat officers in the Army. It is significant, however, that the new tasks required of these officers in the new Army have been identified and analyzed. There is reason to hope that the psychologists, who have been asked to work in cooperation with experienced professional officers on the problem of developing tests for candidates in this area, will provide satisfactory results.

Ever since the development of modern standing armies, maneuvers have been held. These exercises train officers and men in as realistic a way as possible for

combat warfare. Professional military men agree that in maneuvers an evaluation of the performance of individuals and of units may be made with some precision by expert observers. Because no other equally effective means of evaluation seems to exist in peacetime, it is believed that the new test batteries being developed for the assessment of future combat officers may be best judged on the basis of criteria which will include measured performance in realistic maneuvers.

For the past year, the United States Army Scientific Advisory Panel has had a working group of its Human Factors Sub-panel studying the problem of how to improve the selection of future combat leaders. It has already become clear that in attempting to assemble statistically reliable measures in this area professional soldiers and professional psychologists must work together as teams in establishing the new techniques. It is also clear as a result of the preliminary work of the panel that the construction of these new instruments will require a novel combination of previously used tests and of some assessment procedures which have not yet had full statistical validation.

No final report of the work of the Sub-panel has yet been published. It may be said here that the instruments being considered involve the employment of tested psychometric principles and also, in a new way, a rationalization and evaluation of expert opinion.

In summary of the present stage of this research, it can be said that new tactics and new weapons have developed new demands on combat commissioned officers. In the selection and training of such officers modern psychology can be of assistance. This assistance will be of most value if the psychologist works with professional military men. It is a pleasure to report that an effective pattern for this cooperative work has emerged.

Summary

For several months, the authors and a group of colleagues have been studying, from the standpoint of future defense needs, the problem of how best to select army combat officers. The present communication surveys the procedures currently in use and points up areas in which further research is needed.

Particular emphasis is given to the fact that the qualities required of an effective combat officer have undergone marked changes as the result of new developments in weapons and consequent modifications in tactics and the use of personnel. The combat officer of today must possess considerable scientific knowledge and the capacity to form swift, independent judgments, if effective operation of complex modern weapons is to be insured.

In centuries when time was less important, when officer requirements were less specialized, and when the cost of training was lower, officer selection criteria could be defined on a trial-and-error basis. In the larger and vastly more complex modern army, every available scientific method of evaluation must be exploited if satisfactory selection is to be achieved.

Personnel selection and classification methods currently in use in the U. S. Army include the Armed Forces Qualification Test (AFQT) and the Army Classification Battery (ACB). The first of these instruments, developed and used jointly by the American Armed Forces, predicts the degree to which men can be trained for military service. After screening by the AFQT, those individuals accepted for service are administered the ACB, a multi-test instrument measuring seven specific aptitude areas. Use of the ACB to classify men into the military specialties for which they are best suited capitalizes upon special talents, and thus increases the level of job proficiency which can be attained by individuals of given levels of general ability.

Research by Army psychologists on the qualities required of effective combat personnel has resulted in the identification of several measurable characteristics related to combat success in the infantry. Some of these measures, developed partly on the basis of field studies carried out during the Korean War, may also be of value for selecting officers.

The Aptitude for Service Rating developed at West Point Military Academy has been shown to be valid for identifying individuals who are likely to fail in training at officer candidate schools, and job analyses have yielded reliable scales (such as the Aviation Pilot and Navigator Stanines) capable of selecting subordinate officers for special types of duty. At present, however, it is not possible to point to the existence of statistically validated selection instruments for army combat officers.

Progress has been made in identifying the tasks required of combat officers in a modern army, and criteria against which to validate future selection instruments can be obtained in peacetime conditions from observations of performance during simulated war maneuvers. Remaining to be accomplished is the construction of suitable selection instruments and the perfecting of new methods for standardizing and evaluating the observational criteria against which to validate them. This work will demand close cooperation between Army research psychologists and experienced combat officers; it is gratifying that an effective procedure has been designed to facilitate such cooperation.

Résumé

Pendant plusieurs mois, les auters et un groupe de collègues ont étudié, en vue des besoins futurs de la défense nationale, la meilleure façon de sélectionner les officiers d'armée de combat. La présente communication passe en revue les méthodes couramment employées, et signale les domaines dans lesquels il faut pousser davantage la recherche.

On insiste particulièrement sur le fait que les qualités que doit posséder un officer de combat efficient ne sont plus du tout les mêmes qu'autrefois: elles ont été transformées par suite de la création d'armes nouvelles, et les modifications de la tactique et de l'emploi du personnel qui en ont découlé. L'officier de combat moderne doit posséder des connaissances scientifiques considérables, et pouvoir se

former des jugements rapides et personnels, si l'on veut assurer le bon fonctionnement des armes modernes compliquées.

A l'époque où le temps avait moins d'importance, que les habiletés requises des officiers étaient moins spécialisées, et que les frais d'entraînement étaient moins élevés, on pouvait établir des critères pour la sélection d'officiers sur une base empirique. Dans l'armée moderne plus grande et beaucoup plus complexe, il faut exploiter toutes les méthodes scientifiques d'evaluation dont on dispose, si l'on veut arriver à une sélection satisfaisante.

Les méthodes de sélection et de classification du personnel couramment appliquées dans l'Armée des Etats-Unis comprennent l'Armed Forces Qualification Test (AFQT) et l'Army Classification Battery (ACB). Le premier de ces instruments, construit et utilisé conjointement par les forces armées américaines, prévoit le degré auquel les hommes sont capables de profiter de l'entraînement militaire. Après avoir été passés au crible par l'AFQT, ceux qui sont acceptés pour le service sont soumis à l'ACB, une batterie multiple de tests qui mesure sept domaines d'aptitude spécifiques. L'utilisation de l'ACB pour classer les hommes dans les spécialités militaires qui leur conviennent le mieux, exploite des talents spéciaux et augmente donc le niveau de compétence qui peut être attient par des individus de niveaux donnés d'intelligence générale.

La recherche faite par des psychologues de l'Armée quant aux qualités exigées du personnel de combat a permis de déterminer plusieurs charactéristiques mesurables que comporte le succès au combat dans l'infanterie. Certaines de ces mesures, établies en partie en se basant sur des études effectuées sur le théâtre d'opérations pendant la guerre de Corée, peuvent également être intéressantes pour sélectionner les officiers.

La méthode d'évaluation d'aptitude au service mise au point à l'Ecole Militaire de West Point s'est montrée valable pour identifier les individus qui auront le plus de chances d'échouer dans les écoles pour candidats officiers, et des analyses du travail ont donné des échelles sûres (telles que le Stanine-Pilote et le Stanine-Navigateur) permettant de sélectionner des officiers subalternes pour des tâches spéciales. Il n'est cependant pas possible actuellement de signaler l'existence d'instruments de sélection, statistiquement vérifiés, pour sélectionner les officiers d'armée de combat.

Un pas en avant a été fait en ce qui concerne la spécification des tâches requises des officiers de combat dans une armée moderne, et des critères permettant d'évaluer des tests de sélection futurs peuvent être obtenus, en temps de paix, d'après des observations du rendement faites pendant des manoeuvres de guerre simulées. Ce qu'il reste à faire, c'est de construire des instruments de sélection adéquats, et de mettre au point de nouvelles méthodes de standardisation et d'évaluation des critères d'observation du rendement. Ce travail exigera une étroite collaboration entre des chercheurs psychologues de l'Armée et des officiers de combat expérimentés; c'est satisfaisant qu'une méthode efficace a été développée pour faciliter cette collaboration.

Commentary by Discussant

Dr. Wilhelm Witte (Germany):

The proposals of Drs. Carmichael and Baker are so excellent that I can find little to add to their remarks. To a member of a nation in which military psychology ceased to evolve more than a decade ago, it was surprising to learn from this report of the great progress which has been made in the development of officer selection methods.

In Germany, the postwar demilitarisation period has left widespread anti-militaristic attitudes in its wake. As a result, the present role of the military officer in German society is a rather ambiguous one, and this, no doubt, exerts a selective effect on the kinds of individuals who choose military careers. In nations characterized by social attitudes of this kind, a first step in approaching officer selection problems might well be to study the degree of divergence between the actual role of the military officer and popular conceptions of that role. Investigations of this kind, which would throw indirect light on the motivations responsible for self-selection on the part of professional career officers, naturally could not be conducted without prior knowledge of operational requirements. Such information can be obtained through job and task analyses of the type described by the present authors.

Study of officer attitudes and motivations is also important for another reason, namely, that in times of peace it is necessary to select individuals who will become and remain good combat officers when knowing, and even desiring, that they probably will face combat conditions only in simulated maneuvers. Aside from the attitudinal effects on proficiency which are here involved, there is the possibility that selection carried out with due regard for such motivational considerations will eventually produce desired changes in popular stereotypes of the military officer. This, in turn, might be expected to have beneficial effects on the self-selection process. I should therefore like to emphasize here the potential importance of including appropriate measures of attitudes and motivations in combat-officer selection programs.

General Discussion

Dr. Carl J. Lange (U. S. A.):

Some of the work being done at HumRRO in the area of leadership training is relevant to the problems discussed in the preceding paper. One of the duties for which an officer is always responsible is that of orienting the activities of his men toward the achievement of a given goal. The research in question, which dealt with platoon commanders, focused, therefore, on acts directed toward defining group goals and motivating the platoon to achieve them. The analysis was kept at the level of specifics, so that the results could be communicated in meaningful terms to the student officers to be trained.

The procedure consisted of the following steps. First, we collected descriptions of the behavior of platoon leaders in situations involving such activities as describing the job to be done, checking the work accomplished, and reacting to the finished product. For each subject, descriptions were obtained by means of individual interviews with six of the members of his platoon. The obtained pool of behavioral items were then classified into several broad categories (such as "goal-defining behaviors," "motivating behaviors," "information-gathering activities," "handling disruptive influences," etc.), and, subsequently, into more specific subcategories within these broad classes. Objective scores for the categories were obtained by summing across items, and the resulting data were correlated with global performance ratings made by both superiors and subordinates.

Analysis of the results showed that behaviors relating to goal definition, work facilitation, and evaluation of accomplishment were those which correlated most highly with the global estimates of leadership ability. Behaviors which reflected the leader's personality characteristics, such as the degree to which he affiliated with his men, the warmth or coldness of his relations with others, etc., were unrelated to the criterion ratings. It would, appear, therefore, that superiors and subordinates rate leadership ability more in terms of what a man does than in terms of how he does it.

Dr. J. C. Penton (U. K.):

It has always seemed to me that an important shortcoming of attempts to evaluate the personnel and social qualities of prospective officers has been the failure ro relate such characteristics to a general theory of personality. In some work now going on in England, we have been studying the relations between the categories of personality traits used on our officer-selection programs and psychoanalytic theories of personality. I must confess to having been rather unenthusiastic about this particular project when it was first proposed. However, the preliminary results have revealed an astonishing degree of agreement between assessments derived from observations of group performance and evaluations obtained from psychoanalytically-based inventories of likes and dislikes. I shall not go further into the findings here, having merely wanted to call attention to a line of investigation which seems highly promising.

I should also like, at this time, to second the remarks of Dr. Witte concerning the importance of investigating popular conceptions of the military officer. Studies recently completed in England have shown that young men reaching the age of conscription are extremely ignorant of the role of the professional officer, and of the possibilities offered by such a career. As military psychologists, one of our responsibilities is to advise those charged with recruitment about social stereotypes prejudicial to the choice of a military career, and how they can be overcome. Investigations of the type proposed by Dr. Witte can contribute importantly to our accomplishment of this task.

Reply

Dr. Leonard Carmichael:

I should like to thank the speakers for their valuable comments. In emphasizing the need for studying the motivations of career officers and their interrelations with civilian attitudes toward the military, Dr. Witte has drawn attention to some highly important points. Dr. Lange's report on the empirical analysis of leadership behavior is an excellent example of the type of research we have been trying to promote. Finally, Dr. Penton's description of English research on personality assessment also strikes me as being significant. I personally feel that we should consider all possible approaches to the problem of selecting competent officers; I am tempted to say, Mr. Chairman, in view of my prejudices, even the psychoanalytic. But, I am sure this is a matter which would take a very long time to discuss.

A SCORING SYSTEM FOR GROUP INTERACTION IN AN OFFICER-QUALITY ASSESSMENT PROGRAM

Sipke D. Fokkema
Office of Selection Affairs, Royal Dutch Air Force
The Netherlands

There is rather general agreement that a selection program for officer candidates should provide for the measurement of attitudes and social behavior. Among the various techniques for collecting such data, the interview and group observation are perhaps the best known and most widely used.

During the past year and a half, the research group of the Dutch Air Force has devoted considerable effort to improving the traditional group observation method as described by Harria (1949), and as used, for example, by the Assessment Staff of the O.S.S. (1948). In the O.S.S. system, as in the English wartime W.O.S.B. procedure, nearly one-half of the total testing time was devoted to the observation of performance in group task situations. The technique has also constituted an important part of the officer selection program in the Netherlands and in many other countries.

As applied in the above mentioned contexts, the method usually involves the following steps. During the periods scheduled for group observation, the officer candidates, usually in teams of about eight men, are required to carry out various outdoor and indoor tasks. The former may involve such things as crossing a brook with a heavy drum over a bridge improvised by the subjects from a few planks and ropes, or overcoming some obstacle while carrying imaginary explosives. The indoor tasks usually consist of group discussions. In both types of situations the behavior of the candidates is observed by a selection officer who has previously received training in the use of the method and has been warned against such rating errors as halo, overleniency, etc. The officer is provided with a list of personality traits ("confidence," "perseverance," "clamness," "decisiveness," "team spirit," etc.) in terms of which he rates the behavior of the individual candidates. Except for this list, the observer is not guided by any preformulated interpretative scheme, and is left free to make his judgments as he sees fit. In so doing, he usually relies heavily on experience gained from his own assignments in operational or training units.

After the ratings have been obtained, they are submitted to a selection board which holds a conference devoted to joint consideration of each candidate's group performance ratings, objective test scores, and interview results. At the end of the discussion, the chairman of the board makes a final rating of the candidate's overall potential as an officer.

Although the method of group observation has high face validity and, thus, widespread appeal as an officer selection procedure, systematic studies of the technique have raised serious questions as to its predictive value (Eysenck, 1953; Fiske

594869 O - 61 - 10

& Kelley, 1951). Suggested reasons for the uniformly low validities which have been obtained include: (a) low reliability of the ratings, (b) lack of independence among the ratings of different traits, (c) inability on the part of the raters to differentiate behavioral effects due to personality characteristics from those due to socio-psychological factors and specific features of the task situation, and (d) the focusing of attention on personality traits which are, in actuality, unrelated to the criterion to be predicted.

The results of two factor-analytic studies raise further doubts as to the fruitfulness of the traditional procedure. In a study of the dimensions of individual behavior in small experimental groups assigned tasks similar to those used in the W.O.S.B. system, Couch and Carter (1952) identified three main factors: "Individual Prominence" (with an average loading for leadership of .90), "Group Goal Facilitation" (with an average loading for leadership of .35), and "Group Sociability" (with an average loading for leadership of .05). Secondly, an analysis by Halpin and Winer (1952) of the answers given by aircrew members to a descriptive questionnaire on the leadership behavior of airplane commanders yielded the following four factors: "Consideration," "Initiating Structure," "Production Emphasis," and "Sensitivity." These four factors accounted, respectively, for 49 per cent, 33 per cent, 9.8 per cent, and 7 percent of the total variance.

In a recent review, Gibb (1954, pp. 877-920) holds that the factors identified in these and related studies are largely identical, and that the data can be interpreted as support for a three-factor theory of leadership.

Procedure

Collection of Group Observation Data

On the basis of the above results, there were posited at the beginning of the research the existence of three basic dimensions of leadership behavior, called "Sociability," "Effectiveness," and "Individual Prominence." They formed the basic framework of the group task observation system.

Since the research was carried out within an Air Force setting, the set of group tasks usually employed by Army researchers was supplemented with tasks of a technical and administrative nature. It was considered unnecessary to administer several tasks of the same type. Hence, the final battery contained four 50-minute, leaderless-group tasks consisting of one outdoor problem of the usual type, one discussion session, one technical task, and one administrative task.

Following a suggestion by Dr. von Hoesel, a considerably enlarged Poppelreuter test was used as the group technical task. The administrative task consisted of a card-sorting test especially developed for the purposes of the research. The test required that a large number of coded cards be ordered according to a system to be discovered by the subjects. This task, like the Poppelreuter test, involved abstract problem-solving activity.

The study was guided from the beginning by the following principles: (a) the system of scoring should be as objective and reliable as possible, (b) scores for the various tasks should permit objective comparison, and (c) intertrait score contamination should be avoided. For these reasons, the traditional trait-rating procedure was replaced by a modification of the method used by Bales (1950) for recording group interaction processes.

In Bales' system two main aspects of group behavior are distinguished: a "socio-emotional" and a "task" aspect. Each of these is conceived as a dimension having positive and negative poles, both poles being defined in terms of lists of specific behaviors. The task of the group observer is to score each behavior every time it is observed to occur.

The adaptation of the Bales procedure involved the following modifications.[1]

1. Instead of Bales' two dimensions, the three factor-analytically based dimensions specified above were used, modifying the lists of defining behaviors accordingly. Some items from Bales' "task" and "socio-emotional" lists were included in the "Prominence" list, which was extended by the addition of further defining behaviors.

2. The observers were provided with an observation form containing six categories (a "positive" and "negative" category for each of the three dimensions) arranged vertically under the identifying numbers provided for each subject in the group. A model of this form is given in Fig. 1. The observers were instructed to make a mark in the appropriate space for each subject whenever that subject manifested a behavior specified as defining, either positively or negatively, one of the three general dimensions concerned.

3. Orientation of behavior was not scored and time was not recorded.

The same procedure was used for all four group tasks. It yielded comparable data when used by different observers, and provided objective scores which permitted cross-comparison of the subjects' performance on the various tasks.

Identifying No. of Subject		1	2	3
Sociability	(+)			
	(-)			
Effectiveness	(+)			
	(-)			
Prominence	(+)			
	(-)			

Fig. 1.--Model of Group Observation Form

[1]The author is indebted to Lt. Bertholet of the Office of Selection Affairs for his contributions to the development of this method.

Introduction of the new method gave rise to no serious difficulties. Two selection officers were lectured on the operation of the system, and learned the behaviors which were to be counted under the various categories of the observation form. During a practice period of one week, each observer, working under close supervision, scored five groups of subjects on all four of the tasks. After each practice session the scores of the two observers were compared and discussed, every effort being exerted to bring their scoring to a uniform standard.

At the end of the week this aim was largely realized, and the method was used from that point onward for purposes of actual selection. Eleven groups (a total of 87 officer candidates) have since been rated using the new method. Observer A scored 47 and Observer B scored 40 of these subjects.

Collection of Sociometric Data

Also collected for use in the study were sociometric-choice data obtained from the subjects themselves. Since criterion measures could not be expected to become available for some time, it was decided to derive preliminary validity estimates by comparing the scores of the group observers with judgments made by the candidates of each other. Accordingly, at the end of the medical-psychological testing program of three and a half days, each officer candidate was requested to fill out a ten-item sociometric questionnaire designed to cover the three traits rated by the group observers, plus "Leadership" and "Followership." The items of the questionnaire consisted of such questions as: "Who would you most like to have as a friend?", "Who would you most like to have as an assistant?", and their negative variants. For each question the candidate filled in the identifying number of a member of his own test group. Sociometric scores for each of the assumed dimensions were then computed by pooling the choices.

Analysis of the Data and Results

Raw Score Frequency Distributions

A total of 48 raw score frequency distributions was computed (two categories x three dimensions x four tests for each of the two observers). Various comparisons of the medians and ranges of these distributions were made in order to obtain information on interobserver, intercategory, and intertask differences.

1. Comparison of observers. All the raw score distributions were found to be negatively skewed. With only two exceptions, specified below, the ranges of the distributions for the two observers were very similar. The average range of the positive category distributions for Observer A was 0-37; that for Observer B was 0-38. For all but three of the distributions, B's medians were found to be closer to zero than A's.

2. Comparison of categories. The score ranges of the positive categories tended to be somewhat larger than those of the negative categories. For the "Effectiveness" dimension, the positive score range was twice that of the negative.

Comparison of the negative score ranges across tasks revealed marked differences between the observers with respect to rated "Effectiveness" in the technical and administrative tasks. Observer A (a former technician) spread out his negative "Effectiveness" scores for the technical task more than did Observer B.

For the administrative task, on the other hand, the negative "Effectiveness" score range of Observer B (a former code officer) was larger than that of Observer A.

The medians of the negative score distributions were found in all cases to be closer to zero than those of the positive score distributions.

3. Comparison of tasks. No substantial differences were found.

Combining and Transformation of the Positive and Negative Raw Scores

In order to facilitate treatment of the data, each of the 48 raw score frequency distributions was divided into five classes containing, respectively, 10 per cent, 20 per cent, 40 per cent, 20 per cent, and 10 per cent of the total distribution. The resulting ordered classes were then assigned "standard" scores ranging from five to one, with the high positive and low negative ends of the plus and minus distributions being represented by the largest standard scores. For a few distributions, it was necessary to combine classes four and five and assign a value of four and a half.

Contingency coefficients between the rescored positive and negative distributions of each dimension were then computed with respect to all four tasks. This was done for each observer separately, the purpose being to determine whether, for each task and observer, the positive and negative categories of each dimension were sufficiently related to one another to justify addition of the two types of scores.

Although four of the obtained coefficients were found to be very low (i.3., those for the "Effectiveness" dimension on the technical and administrative tasks for Observer A, and those for the "Effectiveness" and "Sociability" dimensions on the technical and outdoor tasks, respectively, for Observer B), the remaining contingency correlations obtained from this analysis were relatively high (.50 - .75). On the whole, therefore, it seemed justifiable to regard the positive and negative categories of each dimension as representing opposite poles of that dimension, and to add up corresponding positive and negative standard scores to obtain a single total score for each dimension on each task.

Interdimension Correlations

After total scores computed as above had been obtained, correlations based on these scores were calculated between the dimensions within each task, and between scores on the individual dimensions across tasks. In computing the interdimension, within-task correlations, the data for the two observers were treated separately for all tasks except the group discussion. For this task, and in computing the across-task correlations for each dimension, the data for the observers were combined.

Selected correlations from among those computed are presented in Table 1. In general, it was found that the correlations between the scores for "Sociability," "Effectiveness," and "Prominence" were both high and similar in magnitude for all four tasks. Moreover, the corresponding interdimension correlations for the two observers were, in general, very similar. In view of the attempts to prevent contamination of scoring through use of the Bales system, and of the checks during the practice week to make sure that the method was being correctly used by the observers, these results came as somewhat of a surprise. The data might be interpreted as indicating that the behavioral dimensions measured were highly task-

TABLE 1
SELECTED CORRELATIONS OF
SOCIABILITY, EFFECTIVENESS AND PROMINENCE DIMENSIONS

A. Between Dimensions: Same Task

Observer	Task	Dimensions Correlated	r	N
A	Outdoor	Prominence--Effectiveness	.79	47
A	Outdoor	Sociability--Effectiveness	.72	47
B	Outdoor	Sociability--Effectiveness	.64	40
A	Technical	Sociability--Prominence	.70	47
B	Technical	Sociability--Prominence	.72	40
B	Technical	Sociability--Effectiveness	.55	40
A+B	Discussion	Sociability--Prominence	.71	87
A+B	Discussion	Sociability--Effectiveness	.57	87
A+B	Discussion	Prominence--Effectiveness	.66	87

B. Between Tasks: Same Dimension

Observer	Dimension	Tasks over Which Correlated	r	N
A+B	Sociability	Outdoor--Discussion	.27	87
A+B	Sociability	Technical--Administrative	.40	87
A+B	Effectiveness	Outdoor--Administrative	.38	87
A+B	Effectiveness	Technical--Discussion	.03	87
A+B	Prominence	Outdoor--Technical	.51	87
A+B	Prominence	Discussion--Administrative	.45	87

bound, and that the various task situations did not differ with respect to them as much as had been assumed.

Comparison of the dimensions across tasks showed that the candidates' judged display of "Sociability," "Effectiveness," and "Prominence" varied markedly from one task to the next. The correlations of each dimension across tasks tended, in general, to be low (see Table 1, Section B, for illustrative data). These results suggest that trustworthy measures of general personality traits or characteristic social behavior cannot be obtained by observing an individual's performance in only a few situations.

Comparison of Group-Observation and Sociometric-Choice Scores

In order to obtain a preliminary estimate of the validity of the group observation procedure, the scores obtained from the selection officers were correlated with the sociometric-choice data obtained as previously described from the subjects themselves. For this analysis, global group-observation scores were computed for each dimension by combining the subjects' standard scores on the four tasks.

TABLE 2
CORRELATIONS BETWEEN GROUP-OBSERVATION AND SOCIOMETRIC-CHOICE SCORES FOR VARIOUS DIMENSIONS

Scores Correlated		
Group-Observation	Sociometric-Choice	Correlation
Prominence	Prominence	$r = .64$
Effectiveness	Effectiveness	$r = .56$
Sociability	Sociability	$r = .14$; $\eta = .42$
Sociability	Followership	$r = .23$; $\eta = .28$

Selected correlations between the global group-observation and the sociometric-choice scores are presented in Table 2. The relatively high (linear) relationships between the observer and peer ratings for "Prominence" and "Effectiveness" suggest that the global group-observation scores computed for these dimensions are psychologically meaningful, and that they may hold promise as criterion predictors.

It will be noted that the relationship between the two types of scores for "Sociability" was somewhat lower and rather markedly curvilinear. The correlation between the observers' scores for "Sociability" and the sociometric-choice scores for "Followership" was very low. A preliminary investigation showed that these two traits were more divergently judged by the candidates than were the other three variables for which peer ratings were obtained. Moreover, the sociograms for "Sociability" and "Followership" appeared to show a greater number of mutual positive and negative choices. In attempting to interpret these findings, one might hypothesize that the buddy ratings reflect more "human" assessments, and that behavior, appearing as highly sociable to the group observers, might have been perceived as ungenuine, exaggerated, or immodest by the candidate's peers.

Summary

An objective method for scoring behavior in group task situations was developed, using a modification of Bales' system for measuring group interaction processes. The technique was found to be practicable and of potential value for the selection of officer candidates. It provides a useful tool for research on group behavior.

Results obtained with the method for a sample of 87 officer candidates underline the necessity of employing group tasks of varying types, if trustworthy estimates of social behavior are to be obtained. The relatively high correlations obtained between group-observer and sociometric-choice scores for "Prominence" and "Effectiveness" indicate that the method yields psychologically meaningful results for these dimensions. Interpretation of the lower, curvilinear relationship between the two types of scores for "Sociability" is a question for further research.

References

Bales, R. F. Interaction process analysis. Cambridge, Mass.: Addison-Wesley, 1950.

Couch, A., & Carter, L. F. A factorial study of the rated behavior of group members. Paper read at Eastern Psychol. Ass., 1952.

Eysenck, H. J. Uses and abuses of psychology. London: Harmondsworth, 1953.

Fiske, D. W., & Kelley, H. H. The prediction of performance in clinical psychology. Ann Arbor, Mich.: Univer. of Michigan Press, 1951.

Gibb, C. A. Leadership. In G. Lindzey (Ed.), Handbook of social psychology (Vol. II). Cambridge, Mass.: Addison-Wesley, 1954.

Halpin, A. W., & Winer, B. J. The leadership behavior of the airplane commander. Columbus, Ohio: Ohio State Univers. Res. Found., 1952.

Harris, H. The group approach to leadership testing. London: Routledge, 1949.

O.S.S. Assessment Staff. Assessment of men. New York: Rinehart, 1948.

Summary

A well-known method of collecting data on attitudes and social behavior for use in officer selection programs is the observation of performance in group task situations. The method involves the rating of various traits displayed by the officer candidates while working in teams on selected indoor and outdoor tasks. Validation studies have repeatedly shown that the rulings thus obtained possess little predictive value. The purpose of the present research was to develop an improved method of group observation.

Several factor-analytic studies have indicated that the variance of leadership behavior in group task situations can be accounted for by as few as three factors. on the basis of these results, the number of traits selected for scoring in the present study was restricted to three: "Sociability," "Effectiveness," and "Individual Prominence." The traditional procedure was further modified by supplementing the set of tasks used by Army researchers with technical and administrative tasks considered to be more representative of Air Force requirements. A considerably enlarged Poppelreuter test constituted the technical task; in the administrative task the candidates were required to sort a large number of coded cards according to a system to be discovered by the group. The complete test program comprised four 50-minute, leaderless-group tasks consisting of one outdoor problem of the usual type, one discussion session, and the technical and administrative tasks described above.

In order to obtain reliable, objective scores permitting of intertask comparison, the usual trait-rating procedure was replaced by a modification of Bales' system for recording group interaction processes. The group observers were provided with a form containing positive and negative categories for each of the three behavioral dimensions to be scored (see Fig. 1 for model), and were asked to learn the lists of specific behaviors to be counted under each category. Each time one of these behaviors was displayed by a candidate during the group sessions, the observers were required to indicate this fact under the candidate's identifying number in the appropriate space on the form.

Two selection officers were trained in the use of the system and, working under close supervision, scored five groups of eight men on all four tasks during a one-week practice period. After each session, the ratings of the two observers were compared and discussed with the aim of bringing their scoring to a uniform standard. This goal being largely achieved, data on a total of 87 officer candidates were collected for research purposes. Observer A scored 47 and Observer B scored 40 of these subjects.

Since criterion data could not be expected to mature for some time, ratings made by the subjects of each other were collected for preliminary validation purposes. The candidates were administered a ten-item sociometric-choice questionnaire dealing with the three variables scored by the group observers, plus "Leadership" and "Followership." Each candidate filled in the identifying number of that member of his own test group he would (a) most prefer, and (b) least prefer in the social roles specified by the items. Sociometric scores for each candidate on the five variables assumed to be measured by the role choices were then computed by pooling the data, and were subsequently correlated with the scores obtained from the group observers.

The analysis of the data and major results can be summarized as follows:

1. All the raw score frequency distributions of the group observers were negatively skewed. The positive category distributions tended, in general, to be characterized by higher medians and wider ranges than the negative category distributions. With only two exceptions, the score ranges of the two observers were found to be highly similar.

2. To facilitate treatment of the data, each of the frequency distributions was divided into five ordered classes containing, respectively, 10 per cent, 20 per cent, 40 per cent, 20 per cent, and 10 per cent of the total distribution. The resulting classes were assigned scores ranging from five to one, the largest values being assigned to the most extreme ends of the positive and negative distributions. Contingency coefficients were then computed between the scores for the positive and negative sides of each dimension for each task, the data for the two observers being treated separately. With the exception of four very low coefficients, the correlations thus obtained ranged from .50 to .75. It was therefore considered justifiable to: (a) regard corresponding positive and negative categories as representing opposite poles of the dimension to which they pertained, and (b) add up their scores to obtain a total score for each dimension on each task.

3. Correlations based on these total scores were then computed between the dimensions for each task, and between scores for the same dimension across tasks. It was found that: (a) the interdimension correlations computed separately for the two observers were very similar in magnitude, (b) the correlations between

dimensions were both high and similar in size for all four tasks, and (c) the subjects' scores on the individual dimensions varied markedly from task to task (see Table 1).

In discussing these results, the author underlines the unexpectedness of the high correlations found between "Sociability," "Effectiveness," and "Prominence" for the various tasks. In view of the efforts made to prevent score contamination, it was anticipated that the ratings would be more independent. The findings may indicate that the dimensions scored were highly task-bound, and that the tasks employed were basically more similar than had been assumed. The low intra-dimension, across-task correlations are interpreted as indicating that performance in many different task situations must be observed, if trustworthy measures of characteristic social behavior are to be obtained.

4. Correlations computed between the sociometric-choice and global group-observation scores (obtained for each dimension by combining scores for the four tasks) showed that "Priminence" and "Effectiveness" tended to be rated similarly by the group observers and the candidates' peers. In contrast, the relation between the two types of scores for "Sociability" was lower and showed a marked tendency to curvilinearity. The correlation between the group-observer ratings for "Sociability" and the peer ratings for "Followership" was very low (see Table 2).

The relatively high observer-peer correlations for the "Prominence" and "Effectiveness" dimensions are regarded as support for the psychological meaningfulness and potential predictive value of the global group-observation scores for these two variables. With respect to the findings for "Sociability," it is suggested that the group observers and the officer candidates tended to perceive such behavior in different ways. Further research on this question is needed.

It is concluded that the method of group observation developed in the study provides a useful tool for research on group behavior, and is of potential value for officer selection purposes.

Résumé

L'observation du comportement pendant l'exécution d'une tâche de groupe constitue une méthode bien connue de rassemblement de données sur les attitudes et réactions sociales des candidats officiers. Cette méthode comporte la valorisation de certains traits manifestés par les candidats officiers s'acquittant en équipe de missions données tant en chambre qu'en plein air. Des études de vérification ont montré a plusieurs reprises le peu de valeur des estimations ainsi établies. Les présentes recherches se sont proposé pour objet de mettre au point une méthode perfectionnée d'observation de groupe.

Diverses analyses factorielles ont indiqué que la variance du comportement de chef dans des missions de groupe peut même s'expliquer par trois facteurs seulement. Sur la base de ces résultats, les traits choisis pour l'affectation de cotes dans le cadre de la présente étude ont été limités à trois dimensions génér-

ales: "Sociabilité," "Efficacité" et "Mise en avant de l'individu." La technique traditionnelle a en outre été modifiée par l'adjonction à la série de tâches utilisées par les chercheurs de l'Armée, de tâches techniques et administratives estimées plus représentatives des exigences de la Force Aérienne. La tâche technique a consisté en un test de Poppelreuter considérablement élargi; la tâche administrative exigeait des candidats le tri d'un grand nombre de cartes codées selon un système que le groupe avait pour mission de découvrir. Le programme complet des tests comportait quatre tâches de groupe sans chef, de 50 minutes chacune, à savoir: un problème en plein air du type courant, une séance de discussion, et les tâches technique et administrative susmentionnées.

Pour qu'elle puisse livrer des cotes objectives et sûres permettant des comparaisons entre les tâches, la méthode habituelle d'évaluation des traits a été remplacée par une modification du système de Bales d'enregistrement des processus d'interaction de groupe. Les observateurs de groupe disposaient d'un formulaire comportant une catégorie positive et une catégorie négative pour chacune des trois dimensions à coter (voir modèle à la Fig. 1). Il leur était demandé d'apprendre les listes de comportements spécifiques à compter dans chaque catégorie. Chaque fois qu'un candidat donné manifestait un de ces comportements au cours des séances, l'observateur l'a indiqué sous le numéro de référence du sujet dans l'espace prévu à cet effet sur le formulaire.

Deux officiers sélectionneurs formés à ce système, travaillant sous étroit contrôle, ont coté cinq groupes de huit hommes pour chacune des quatre tâches au cours d'une période d'application d'une semaine. Après chaque séance, les résultats des observateurs étaient comparés et discutés pour rapprocher leurs méthodes de valorisation. Ce but largement attient, des données pour un total de 87 candidats officiers ont été recueillies aux fins de la recherche. Les observateurs A et B ont coté respectivement 47 et 40 de ces sujets.

Des critères opérationnels ne pouvant pas être obtenus avant quelque temps, des estimations faites par les sujets l'un de l'autre furent recueillies aux fins de validation préliminaire. Un questionnaire sociométrique de dix items portant sur les trois variables cotées par les observateurs de groupe complétées par "Aptitude de chef" et "Aptitude de subordonné," fut distribué aux sujets. Chaque candidat a donné le numéro de référence de ce membre de son propre groupe de test qu'il eût (a) le plus et (b) le moins préféré voir dans les rôles sociaux précisés par les divers items. Les cotes sociométriques pour chacune des cinq variables ainsi mesurées furent ensuite corrélées avec celles livrées par les observateurs de groupe.

L'analyse des données et les résultats principaux peuvent se résumer comme suit:

1. Toutes les distributions de fréquence des cotes brutes des observateurs de groupe furent négativement déviées. En général, les distributions de catégorie positive tendaient à se caractériser par des valeurs médianes plus élevées et des écarts plus larges que les distributions de catégorie négative. A deux seules exceptions près, les écarts des distributions des deux observateurs ont été très semblables.

2. Pour faciliter l'analyse des données, chacune des distributions de fréquence des cotes brutes fut divisée en cinq classes ordonnées comportant respectivement 10, 20, 40, 20 et 10 pour cent de la distribution totale. Chaque classe était affectée d'une cote allant de cinq à un, les valeurs supérieures étant données

aux extrémités des distributions positive et négative. Des coefficients de contingence basés sur ces cotes furent alors calculés entre les aspects positif et négatif de chacune des dimensions pour chacune des tâches, les données des deux observateurs étant traitées séparément. A l'exception de quatre coefficients très bas, les corrélations ainsi obtenues allaient de 0,50 à 0.75. On s'est donc jugé autorisé à: (a) considérer les catégories positive et négative appartenent à la même dimension comme en représentant les pôles opposés, et (b) additionner les cotes positive et négative pour obtenir des cotes totales pour les dimensions.

3. Des corrélations basées sur ces cotes totales furent alors calculées entre les dimensions pour les quatre tâches et, pour chacune des dimensions, entre les diverses tâches. Il apparut que (a) les correlations entre dimensions calculées séparément pour les deux observateurs étaient tres voisines en grandeur, (b) les corrélations entre dimensions étaient à la fois élevées et de grandeur voisine pour les quatre tâches et (c) les cotes des sujets pour chaque dimension variaient nettement d'une tâche à l'autre (voir Table 1).

Dans sa discussion de ces résultats l'auteur souligne le caractère inattendu des étroites corrélations constatées entre "Sociabilité," "Efficacité" et "Mise en avant" pour les diverses tâches. Eu égard aux précautions prises pour éviter la contamination des cotes, on était en droit de s'attendre à des estimations plus indépendantes les unes des autres. Les résultats obtenus peuvent être interprêtés comme une indication que les dimensions cotées étaient sous l'étroite dépendance des tâches et que celles-ci étaient à la base plus apparentées qu'on avait pu au début le supposer. La constatation de faibles corrélations entre les valeurs d'une même dimension pour les diverses tâches passe pour indiquer que, si l'on veut obtenir des mesures fidèles du comportement social, il faut observer celui-ci dans le cadre de situations nombreuses et variées.

4. Les corrélations calculées entre les cotes sociométriques et les cotes globales d'observation (obtenues pour chaque dimension par combinaison des cotes afférentes aux quatre tâches) ont indiqué que "Mise en avant" et "Efficacité" étaient estimées semblablement par les observateurs de groupe et les candidats eux-mêmes. Par contre, la relation entre les deux types de cotes afférentes à la "Sociabilité" était moins marquée et manifestait une tendance nette à la curvilinéarité. La corrélation entre les estimations des observateurs de groupe pour la Sociabilité" et celles des candidats pour l'"Aptitude de subordonné" s'est révélée très faible (voir Table 2).

Les corrélations relativement étroites entre les résultats des observateurs et ceux des candidats pour les dimensions de "Mise en avant" et d'"Efficacité" s'interprètent comme étayant la signification psychologique et la valeur potentielle de prédiction des cotes globales de l'observation calculées pour ces deux variables. En ce qui concerne les constatations pour la "Sociabilité," on peut estimer que les observateurs de groupe et les candidats officiers ont eu tendance à concevoir ce comportement sous des angles différents. Cette question nécessite des recherches plus approfondies.

On peut conclure que la méthode d'observation de groupe mise au point dans le cadre de cette étude fournit un utile instrument de recherche sur le comportement de groupe et le problème de la sélection des officiers.

Commentary by Discussant

Dr. Abraham Carp (U. S. A.):

The use of situationsl tests for the assessment of personality and leadership characteristics has had a long and checkered history. Because of its "real life" qualities, the procedure has appealed to military personnel who sometimes find it difficult to accept the more esoteric tools of the psychologist.

One great drawback of the method has been the difficulty of objectively demonstrating its validity. The Germans, British, and Americans used situational tests in the selection of officers and special-duty personnel during World War II, but follow-up was difficult and never satisfactorily complete. Since the war, numerous validation studies have been carried out in the United States with essentially discouraging results. An extensive review of the literature is given by Cronbach in the 1956 Annual Review of Psychology; except for buddy rating techniques and group discussion procedures, his evaluations are generally negative.

Dr. Fokkema's study represents an attempt to provide a theoretical rationale for situational test scoring and a more objective method of recording behavior in group task situations. It is unfortunate that his design did not provide for having both observers rate the same individuals, at least for a sample of the subjects, so that a better measure of inter-rater agreement than similarity of score distributions would have been available. Other studies, however, have indicated that fairly high agreement can be obtained by use of the Bales system, so, perhaps, this is not too critical a point.

Although the correlations between the dimensions were found to be relatively high, it is still possible that a factor analysis of the correlation matrix would, after removal of the general factor, reveal the original three-factor structure which was introduced. The correlations found in the study between scores for the same dimension across tasks indicate a good deal of situational specificity. This is a phenomenon frequently encountered in the field of assessment.

I believe that, as evidence for the validity of the observation procedure, the correlations between the group-observer and sociometric ratings should be viewed with extreme caution. First of all, the two types of ratings were virtually concurrent, so that the long-range validity of the group-observer scores remains unknown. Second, because of this short time lapse between the two types of ratings, it is probable that the various random errors of measurement were not independent. It might have been interesting to compute correlations among the peer ratings, so that the interdimension relationships characterizing these data could have been compared with those based on the group observation results.

Recent studies in the United States have raised some doubts about the meaningfulness of the peer-rating criterion, which is so often used in validating various types of predictors. Situational specificity and the particular nature of the group have been found to play a larger role than is generally assumed. For example, some data recently analyzed in our laboratory showed that peer choices, made in similar situations by the same subjects with respect to "most likely to make a good officer," correlated .90 with each other when there was a five-week period separating the two sets of ratings. After a three-month interval, however, when the situation had changed from preflight training to primary flight conditions, and the nature

of the group was accordingly altered, the correlation between ratings dropped to .57. Similar findings have been reported by Navy researchers. Thus, while it is true that peer ratings tend to have high correlations with later performance measures (which, in the military situation, uaually are also ratings) we still do not have a complete understanding of these measures. Some of our data indicate that the peer is not necessarily a better judge of future performance than, say, the instructor, and that it is merely the larger available number and, hence, higher reliability of peer ratings which is responsible for their apparently higher validity.

Several studies of situational tests are currently underway in our laboratory. In one intensive investigation, based on a sample of 270 officer candidates, a very wide variety of tasks was employed. Among them were six different role-playing tasks. Two observers independently: (a) scored the subjects on a number of specific, objectively observable behaviors, and (b) rated them on 30 personality and temperament traits. For the objective check list, inter-rater reliability was found to be about .75; for the trait ratings, about .25. Significant in relation to Dr. Fokkema's results was the finding that the interscore correlations across situations were very low, and that none of the clusters (developed by applying homogeneous-keying techniques to the objective check-list data) contained items from more than one situation. Again, therefore, we find evidence for considerable situational specificity in assessment procedures of this type.

Data such as these support Cronbach's contention that, despite their length and apparent internal diversity, situational tasks should be considered essentially as test items. It is certainly true that the correlations usually obtained across situations are of the same order of magnitude as those between test items. It is my personal belief that we should adhere in this field to the traditional psychometric approach, and employ a wide variety of situations in expectation of obtaining a general score or set of scores having higher validities than the individual tasks. Such an approach would require that large numbers of scores and tasks be tried, and the best ones picked out and refined by means of the procedures used for item selection and validation. More rapid progress would probably be made if the original scores and situations were to be selected in terms of a theoretical rationale, as Dr. Fokkeme has done, rather than on a strictly empirical basis.

General Discussion

Dr. Simon Biesheuvel (Union of South Africa):

Group situational test procedures seem to have taken a bit of a beating this morning, and I should like to say a few words in their defense. We have made extensive use of leaderless group tasks in selecting African gang leaders for the gold-mining industry. After having been screened on general-ability performance tests, the candidates participate in a series of group tests similar to those employed in the British services for purposes of officer selection. Validation of the results against success in training has yielded criterion correlations of around .60 for numerous samples, and reasonable validities have also been obtained against the criterion of achievement on the job. Thus, whereas the wastage in the training schools

used to be in the vicinity of 26 per cent, it has now been reduced to approximately 3 per cent. These procedures have been used in the Union of South Africa for about six years, and have been extended to West Africa, Kenya, and the copper belt in Rhodesia.

I should also like to mention some research dealing with the selection of candidates to be trained for administrative work in a large industrial corporation. The methods used were similar to those originally developed by the British Civil Service Selection Board. The validation study is of particular interest because the corporation in question insisted that all its own employees, including those who did not possess the academic qualifications set for outside candidates, be allowed to apply for training. The result was that a large number of company employees went through the testing program and, upon failing to be selected, returned to their former jobs. We were therefore able to follow up not only those who were selected for training but also those who were not selected, and, by developing a special assessment scheme for the latter group, it was possible to collect appropriate criterion data. The validity coefficients obtained under these conditions were over .60, and the method has now been in routine use by the corporation for at least six years. In our experience, therefore, group-task procedures have been found to be of considerable value for purposes of selection in widely differing circumstances.

Reply

Dr. Sipke D. Fokkema:

With regard to Dr. Carp's remarks on the value of peer ratings as criterion measures, I should like to emphasize that our use of such ratings was conceived only as a preliminary step. Since these sociometric data were collected at the same time as the group observation records, they obviously cannot be considered as a criterion of job or training success.

It should also be mentioned that we did compute the intercorrelations of the sociometric scores, and found some interesting differences between these coefficients and those obtained for the group observation data. I won't go into these results for fear of getting involved in a separate communication on peer ratings. Let me conclude, therefore, by saying that we hope to continue our work with group test methods and to look into a number of as yet unexplored problems—including, of course, the question of the value of such methods for purposes of prediction.

Concerning Dr. Biesheuvel's remarks, there is naturally a great difference between the situations in which Air Force officers and gang leaders for a gold mine have to be selected. There are important educational and economic background differences, and it has been my experience that these have a great deal to do with the validities found.

MANAGEMENT TRAINING FOR NON-COMMISSIONED OFFICERS

P. J. Sadler
The Air Ministry
London, England

In a modern, highly technical fighting service, such as the Royal Air Force, the task of the regular Senior NCO is a difficult and exacting one. Besides being expert in his particular trade, he must be able to plan the work of his section, organize work teams, control the quantity and quality of output, obtain maximum utilization of manpower and equipment, and so on. Not least, he must be a skilled practitioner in human relations. The average NCO, in the past, has had little opportunity to acquire the knowledge and skills necessary to fit him for the supervisory, as distinct from the technical, aspects of his job.

The Need for Management Training

In recent years, the inadequacy of NCO training and development in these respects has been demonstrated in the course of various investigations into efficiency and morale. Work-study teams have surveyed a number of R.A.F. establishments and have shown how manpower, equipment, and other resources can be saved by the application of simple method-study techniques. On the human relations side, inquiries into internal recruiting have shown that many servicemen cite relations with their NCO's as an important factor deterring them from signing on. During the course of a field study of factors influencing airmen's attitudes, Air Ministry psychologists had an opportunity to investigate problems of NCO-airmen relations at first hand. It was found that the attitudes towards NCO leadership expressed by airmen on the station being studied differed quits radically from one section to another, and that these differences were closely related to measured differences in morale. Those who expressed unfavorable attitudes mentioned, among other factors, such shortcomings on the part of NCO's as failure to delegate authority and responsibility, "over-supervision," lack of job-helpfulness, failure to explain orders, and lack of tact in dealing with subordinates. Similarly, interviews with Senior NCO's revealed sharp differences among them in such matters as attitudes to subordinates, and leadership philosophy in general.

In its concern over these problems the R.A.F. has turned to industry for help on how it can improve the productivity and human relations skills of its personnel. As a consequence, it has tended to become "management conscious" rather than "leadership conscious," a phenomenon which has in itself helped to get the program accepted. For, although Air Force officers are willing to concede that psychologists can give them valuable advice on "management," they are much less willing to accept advice on "leadership," which they regard as their own particular province.

Requirements of a Management-Training Program

Any course of training designed to improve the management and human relations skills of Senior NCO's must fulfill a number of basic requirements if it is to achieve success.

594869 O - 61 - 11

First, it should be related to the particular conditions of a military service, and should be designed specifically to meet the needs of the service NCO. (It is not a good substitute to send NCO's to civilian courses in foremanship.) This requirement, however, raises certain difficulties. The facts, cases, and experiments, which form the basis of theories in management and human relations, are very largely derived from industrial settings. NCO's find it difficult to accept, as applying to the service, principles based on industrial experience, however broad and objectively recorded that experience may be. They prefer to be guided by their own particular, subjective experience of life in the service. Until such time as the usefulness of modern management techniques can be effectively demonstrated within the service itself, there will be reservations of this kind in the minds of trainees. Part of any management training program should involve the building-up of illustrative experimental material from existing service sources and the development of new methods of generalizing and objectifying that experience.

A further difficulty which presents itself at this stage is the development of criteria of effectiveness appropriate to military groups. In industry, the criterion against which management techniques are evaluated is usually one of productivity. In most military groups, however, direct measurement of productivity is impossible. Various other indicators of efficiency must be resorted to—indicators which are usually less reliable and, at the same time, less convincing.

A second requirement of an NCO management training program is that it must be aimed at the right level. In other words, it should be designed to fit the trainee to handle those problems in management and human relations which commonly occur at the NCO level. It should not attempt to cover principles and functions appropriate to the higher executive and administrative levels. This is stating the obvious, but, since programs of this nature are initiated and drawn up by people accustomed to thinking in terms of administration rather than of supervision, it cannot be overstressed.

A third requirement is that there should be a strong incentive for the NCO to learn as much as he can from the course. Such an incentive could perhaps be provided by making successful completion of the course an essential qualification for promotion. This would at least insure that NCO's will want to take the training. In addition, the quality of the instruction should be such that it "sells" itself to the trainee, and makes him want to take advantage of the opportunities that it offers. He should also be able to feel that his superior officers will not stand in the way of any improvements he might wish to make as a result of his training.

Whether or not NCO's can be "solc" on management training depends largely, of course, on the quality of instruction. These subjects cannot be taught "from the book." Management instructors need to be of very high quality; they need to be experienced, widely read, fluent speakers with skill in personal relationships. These qualities are unlikely to be found among Senior NCO's; teaching management should undoubtedly be a job for an officer.

Management Training in the Royal Air Force

Management training for Senior NCO's in most trade groups in the Royal Air Force has now been in process for more than a year. A separate course is conducted for each trade group. Each course is of eight weeks' duration and consists of two parts: five weeks of trade "refresher" training, and three weeks of management. The management training is further divided into three phases as follows:

1. Principles and functions of management and human relations (23 hours). This phase is designed: (a) to provide the NCO with a better knowledge and understanding of the principles and functions of management and their applications in the armed services; (b) to acquaint the NCO (in most cases, for the first time) with the findings of relevant research in the spheres of military and industrial psychology, and human relations; (c) to break down stereotyped, prejudiced ways of thinking and replace them with more realistic attitudes toward personnel problems; (d) to improve skills in handling people; and (e) to give the NCO confidence in his powers of leadership, and thus enable him to translate what he has learned from the course into practice.

2. Work study (23 hours). This phase provides an introduction to the fundamental principles and basic techniques of work study, and is intended: (a) to enable the NCO to apply simple work-study techniques to his own section in order to obtain greater efficiency, and (b) to increase his understanding of the aims and methods of work study, and thus insure his active cooperation with visiting work-study specialist teams.

3. Technical organization and administration. This portion of the course is aimed at rendering the NCO thoroughly familiar with the machinery of administration in his trade, i.e., with such matters as service writing, office management, equipment procedures, technical records, statistics, and so on.

Successful completion of the course has now been made a condition for promotion to the ranks of Flight Sergeant and Warrant Officer. Furthermore, the R.A.F. gives so much importance to this type of training that it has been decided to add bonus points to NCO promotion scores according to the level at which the course was passed. The number of bonus points allotted are three for a "distinction" (A.1) pass, two for a "very good" (A.2) pass, and one for a credit (B.2) pass.

Nature of Management and Human Relations Training

Air Ministry and Technical Training Command psychologists were given the task of drawing up a syllabus for the first 23 hours of management training ("Principles and Functions of Management and Human Relations"). Obviously, 23 hours is not a great deal of time for a field of this magnitude. Accordingly, it was decided to concentrate on essentials only, and to give these as full a treatment as possible in the time available. It was further decided to center the phase around the main responsibilities of the NCO—the maintenance of efficiency, morale, and discipline within his section—and to show how NCO behavior can affect these various aspects of organizational efficiency and the relations between them. The syllabus finally recommended and adopted was as follows:

1. Introduction to management phase of the course	Lecture	1 hour
2. Introduction to principles and functions of management	Lecture	1 hour
3. The application of management principles in the services	Lecture	1 hour
4. The Senior NCO and efficiency	Lecture	1 hour

5.	Effective instruction	Practicum	1 hour
6.	How efficiency can be improved under present conditions	Discussion	1 hour
7.	Case studies in efficiency	Syndicated task	3 hours
8.	What is morale?	Lecture	1 hour
9.	Morale and leadership	Lecture and Discussion	2 hours
10.	Morale and efficiency	Lecture	1 hour
11.	Morale and discipline	Lecture	1 hour
12.	Case studies in morale	Syndicated task	3 hours
13.	Giving orders	Practicum	2 hours
14.	Administering praise and blame	Practicum	2 hours
15.	Course summary	Lecture	1 hour
16.	Final discussion	Discussion	1 hour
		Total:	23 hours
		Lectures:	9 hours
		Participatory forms of training:	14 hours

As can readily be seen from the syllabus, very full use was made of participatory forms of training, e.g., case studies, group discussions, role-playing. A number of case studies were prepared in advance, and instructors have subsequently been encouraged to develop their own. Each case involves a relatively simple supervisory problem of the kind which the NCO in charge of a section is commonly called upon to resolve without reference to his superior officer, and which can easily be related to the principles introduced in the lectures. The trainees discuss each case presented to them and attempt to reach a workable and satisfactory solution under the guidance of their instructor, who acts as discussion leader. There are no "official" solutions to these cases. However, the group leader must see that the discussion keeps to the facts, that the realities of the situation are respected, and that the problems are related to the principles involved in them. Once a generally acceptable solution has been reached, a short period of role-playing ensues. In this, the trainees attempt to apply, in face-to-face situations, the courses of action they have recommended.

The two great virtues of these methods are that: (a) their novelty and intrinsic interest value keep the NCO's actively participating in the training, and (b) group pressures, which are so effective in producing attitude changes, are very considerably utilized.

Use is also made of films and other types of visual aids. A recent innovation has been the presenting of case studies in the form of film strips. This method of presentation seems to facilitate conceptualization of the problem situations, makint it easier to concentrate on essentials in the discussions. In addition, the situation is made through dramatization to seem more real, enabling the trainees more easily to get the feel of it.

A further training aid employed is the "management notebook," a small booklet containing skeleton outlines of the lectures, together with space for the NCO's own notes. This is intended to encourage intelligent notetaking, and also to serve as a "supervisor's manual" when the NCO returns to his unit.

Evaluation of the Management Training

Subjective impressions obtained from visits to the courses indicated that the training was achieving considerable success in improving the attitudes of NCO's. It was planned at the outset, however, to attempt a more objective evaluation of the merits of the training. This evaluation will be made at two levels. At the attitude level, changes are to be measured by means of questionnaires administered before and after training. At the behavioral level, it is intended to follow up trainees who have returned to their units so as to determine whether any improvement in their supervisory behavior has taken place. Various criteria—such as ratings by superior officers, measures of the morale of subordinates, and, where applicable, productivity indices—will be employed in this phase of the research.

A study of attitudes is now almost complete, and the indications are that significant improvements in NCO attitudes can result from the course. For a sample of a hundred trainees, it was found that the group as a whole showed a highly significant change in attitude scores at the end of the training. When the sample was broken down into individual trade groups, however, it was found that groups having high intelligence and educational levels scored initially higher on the attitude questionnaire, and tended to improve much more than did groups having low intellectual and educational requirements. Those trade groups which had the lowest pre-training scores, e.g., cooks and telegraphists, failed to show any significant improvement as a result of their training. It would therefore appear that certain conditions are necessary if instruction of the type described is to take effect.

Conclusion

As the technical complexity of modern fighting services increases, the need for training in managerial skills and techniques becomes more and more evident. The beginning made by the Royal Air Force in this area has been described, but not with the intention of holding up these training methods as an ideal, or even as an example to be followed. The main purpose of the paper is to stimulate discussion regarding the optimal form of a management training course designed specifically for the armed forces.

Summary

The communication describes the nature and purpose of the NCO management training program recently instituted by the Royal Air Force. In addition to being expert in the technical aspects of his work, the Senior NCO should be skilled in human relations and capable of planning, organizing, and supervising the work of his section so as to achieve maximum utilization of manpower and equipment. In the past, training in these important management skills has been inadequate, with consequent deleterious effects on Air Force efficiency and morale. Research has shown, for example, that the quality of NCO-airman relations is closely related to measured differences in morale, and enters as a factor into decisions to re-enlist.

An effective management training course for non-commissioned officers should fulfill the following requirements: (a) it should be bound to the military situation and focused on the specific needs of the service NCO; (b) it should be aimed at the NCO level and not attempt to cover problems germane to higher executive and administrative positions; (c) strong incentives to learning should be provided; (d) the instruction should be interesting and efficient; and (e) the NCO should feel that he subsequently will be free to put his newly learned principles into practice. In connection with point (c), it is noted that in the Royal Air Force successful completion of the course has been made a qualification for promotion, and bonus promotion points are assigned in proportion to success in the course as reflected in final grades. As to point (d), it is held that management training should be taught by officers, if high-quality instruction is to be measured.

Management training for Senior NCO's in most R.A.F. trade groups has been in progress for more than a year. A separate course is conducted for each trade group. All courses consist of five weeks of "refresher" technical training, followed by three weeks of instruction in the techniques of management, the principles of human relations, methods of work study, and problems of organization and administration. Emphasis in the "management" portion of the course is on the understanding of basic principles, the communication of relevant research findings, the changing of stereotyped attitudes, and the improvement of skill in the handling

After specifying the course content in greater detail, the training aids utilized are described. Considerable use is made of participant forms of learning, such as role-playing and group discussion of case studies. Films and other visual aids are also employed.

Plans for evaluating the effectiveness of the training include: (a) administering attitude questionnaires before and after the course, and (b) following up the NCO's after return to their units to determine whether any improvement in practical supervisory behavior has taken place. Ratings by superior officers, measures of the morale of subordinates, and, where possible, indices of productivity will be employed to measure changes in on-the-job effectiveness.

In a study already completed, a significant improvement in measured attitudes was found for a group of a hundred trainees upon completion of the course. When the sample was broken down into individual trade groups, however, it was found that the most intelligent and educated trainees both: (a) obtained higher scores on the attitude questionnaire prior to training, and (b) showed the greatest amount of improvement as a result of the course. The least intelligent and educated NCO's had initially poorer attitudes, and showed no significant improvement

at the end of training. It appears that certain conditions must be present if the training is to have any effect.

The author invites discussion on the question of how best to design courses in management training within a military context.

Résumé

La communication décrit la nature et le but du programme de "management training" (formation administrative) récemment institué par la Royal Air Force à l'intention de ses sous-officiers. Outre qu'il doit posséder son métier, le sous-officier doit être habile en ce qui concerne les relations humaines, et être capable de projeter, d'organiser et de contrôler le travail de ses subalternes de façon à obtenir le maximum de rendement. Jusqu'à présent, la formation dans ces importantes compétences de direction a été insuffisante, ce qui a eu pour conséquence de nuire au rendement et au moral. La recherche a montré, par exemple, que la qualité des relations entre les sous-officiers et leurs subalternes est en relation étroite avec des différences mesurées du moral, et qu'elle intervient, pour ces derniers, dans leur décision de se réengager.

Un cours efficace de formation de direction pour les sous-officiers doit remplir les conditions suivantes: (a) il doit être lié a la situation militaire et centré sur les besoins spécifiques du sous-officier; (b) il doit rester au niveau du sous-officier et ne pas essayer d'englober des problèmes se rapportant à des postes supérieurs; (c) des stimulants puissants à l'étude doivent être assurés; (d) l'enseignement doit être intéressant et efficace; et (e) le sous-officier doit sentir qu'il sera libre après de mettre en pratique ses connaissances nouvelles. En ce qui concerne le point (c), on note que, dans la Royal Air Force (Armée de l'aire Royale), la réussite au cours est devenue une condition de l'avancement, et des points supplémentaires pour l'avancement sont accordés en fonction du niveau de la réussite au cours. Quant au point (d), on estime qu'un enseignement valable ne peut être assuré que par des officiers.

La formation administrative pour des sous-officiers appartenant aux divers groupes techniques de la R.A.F. s'est développée depuis plus d'un an. Un cours séparé est donné pour chaque spécialité. Tous les cours comportent cinq semaines d'enseignement technique à titre de révision, et trois semaines d'enseignement relatif à des techniques de direction, des principles de relations humaines, des méthodes de l'étude du travail, et des problèmes d'organisation et d'administration. Dans la partie du cours portant sur le "management," on met l'accent sur l'explication des principes de base, la communication de constatations de recherche pertinente, la modification des attitudes stéréotypées, et l'amélioration des compétences sociales.

Après avoir détaillé le programme du cours, les méthodes d'enseignement utilisees sont décrit. On fait un usage étundu de méthodes faisant participer personnellement l'élève, telles que l'interprétation d'une rôle social et la discussion générale des problèmes représentatifs. Des films et d'autres méthodes visuelles sont également employés.

Pour évaluer l'efficacité du cours, on a prévu: (a) de donner, avant et après le cours, des questionnaires portant sur les attitudes; (b) de suivre les sous-officiers après le retour dans leurs unités pour déterminer si le fonctionnement en ce qui concerne la direction a été amélioré. Des évaluations faites par des officiers supérieurs, des mesures du moral des subordonnés et, si possible, des indices de productivité, seront utilisés pour mesurer des changements dans l'efficacité pratique.

Dans une étude déjà complétée sur un groupe de cent sujets, on a constaté une amélioration importante dans les attitudes mesurées à la fin du cours. Cependant, quant l'échantillon a été divisé en groupes techniques individuels, on a trouvé que les sujets les plus intelligents et les plus instruits: (a) ont possédé de meilleures attitudes avant de commencer la formation, et (b) ont fait preuve d'une plus grande amélioration par la suite. Les sous-officiers les moins intelligents et instruits avaient de plus mauvaises attitudes au début, et n'ont montré aucune amélioration significative à la fin. Il semble qu'il faille certaines conditions préalables pour que la formation ait de l'effet.

L'auteur invite des suggestions portant sur la meilleure façon de concevoir des cours de direction pour les chefs militaires.

Commentary by Discussant

Dr. Kullervo Rainio (Finland):

It seems to me that the course outlined by Mr. Sadler, with its emphasis on efficiency, morale, and discipline problems, constitutes a most fruitful starting point for NCO supervisory training. I was also impressed by his thorough plans for evaluating the effects of the instruction. I wonder, however, whether much change in the attitudes and productivity of subordinates will occur so long as the group within which the NCO works remains the same after his training as before it. The attitudes of subordinates toward their leader are strongly affected by their former experience with him; hence, the task of the retrained NCO may be rendered more difficult by the necessity of coping with previously developed expectations. If possible, it might be desirable to assign some of the NCO's to new units after their training, so that the results can be compared with those obtained in units which receive their own NCO's back again.

One other minor point concerns the use of instructors as discussion leaders in the case-history sessions. It has been our experience that the amount of active participation and the effects of group pressure tend to be greater if discussions are led by a group member rather than by an "official."

A more important question concerns the problem of coordinating the NCO training with that of superior officers. How can one insure that the attempts of an NCO to put his newly acquired management principles into practice will be accepted and endorsed by his superiors? This problem calls to mind the case of the industrial supervisor who passed a leadership course in our Institute with brilliant suc-

cess. Full of enthusiasm, he went directly from the course to the company president, saying, "This company is thoroughly rocky. Everything must be changed." The only change that resulted, of course, was one of personnel: the supervisor was ejected from the company.

Another point I should like to mention is the possibility of incorporating the so-called "feed-back" method into the training program. Research has shown that one of the most effective ways of increasing productivity and inducing positive attitudes toward management in industrial situations is to make use of a circular study-action procedure. For example, a department is first surveyed by means of an attitude questionnaire. Then, the results are discussed jointly by supervisors and subordinates, decisions are made on a group basis, and the changes decided upon are introduced. After a suitable lapse of time, the attitude questionnaire is re-administered and the whole process is repeated. Such a technique might well be worth trying in military groups, where problems of morale and motivation are of particular importance.

A final question concerns the method used to provide incentives for participants in the training course. The practice of making this short course a condition for promotion, and of assigning promotion points according to the degree of success, fills me with some concern. Many studies have shown that intellectual ability is not the most essential factor in leadership. The correlations between general intelligence and success in industrial supervisory positions are usually found to be low; around .25 to .30. In a study of my own on industrial leadership, the regression of supervisory success on intelligence was significantly non-linear, with maximum success occurring about one sigma above the mean of the IQ distribution. Of the three independent factors found in this study to be correlated with supervisory success, intelligence was less important than the other two: non-suggestibility and emotional stability. It seems reasonable to assume that success in the training course described by Mr. Sadler depends rather heavily on general intelligence. I would, therefore, tend to question whether screening the NCO's on the basis of their course grades will result in optimal selection for purposes of promotion.

General Discussion

Col. Carlson (U. S. A.):

In the second section of Mr. Sadler's paper, it is stated that "teaching management should undoubtedly be a job for an officer." In our Air Force NCO academies, we invariably use NCO's as instructors, the source generally being the most successful graduates of the academies. The results have always been satisfactory. Mr. Sadler also states that one of the objectives of the R.A.F. training course is to change attitudes. It seems to me that greater success might be achieved here if men, who are themselves NCO's, and thus familiar with the particular problems they face, were to be used as instructors.

Reply

Mr. P. J. Sadler:

Dr. Rainio's question about how to insure coordination between the training given to NCO's and that given to superior officers is a very vexing one for us. At the moment, we are in the strange position of feeling that the leadership training received by NCO's is of much better quality than that received by their superiors. Others are also becoming aware of the discrepancy, however, so we hope soon to be given responsibility for developing training programs for superior officers as well. If and when this occurs, the officer training will, of course, be planned so as to complement that provided for the NCO's.

I should now like to touch on the question of whether or not it is desirable to make the NCO training course a condition for promotion. We feel that it is desirable—primarily, because to date we have been unable to obtain satisfactory evaluations of suitability for promotion from commanding officers. For example, when we ask officers to rate their NCO's on a five-point scale, it invariably happens that about 90 per cent of the rates are placed in the first category. Consequently, it was felt that advantage should be taken of the assessment opportunitye presented when NCO's from various units were brought together at a central training school.

As to Col. Carlson's suggestion, using NCO's rather than officers as instructors would present some difficulties in our particular situation. First of all, most of the Senior NCO's who would be eligible to teach are old-line men who failed to achieve the rank of officer even at a time (during the war) when the officer shortage was so severe that the R. A. F. was commissioning virtually every NCO who could conceivably do the job. Thus, our present cadre of NCO's consists, to put it bluntly, of left-overs, and we cannot hope to find among them many individuals who possess the qualities desired in an instructor. Moreover, even if we were to find some good instructor material among these NCO's, we do not have the authority to decide which men are to be assigned to teaching duty. Such decisions are made by the administrative branches in terms of other criteria.

As to the role played by the instructor in changing attitudes, it has been our experience that the case-study discussion method contributes more in this respect than the personal qualities of the instructor. Our technique has been to have the NCO's write individual solutions to the problems presented by the case stucies, subsequently to discuss the case among themselves in six-men groups, and then to compare the group with the individual solutions. The trainees were often so amazed at the extent to which their outlook had altered as a result of the discussions that they tended to regard the technique as responsible for the change, this being true regardless of whether our previous assessments of the ability of the instructor had or had not been favorable.

CONTRIBUTION OF THE INTERVIEW TO PREDICTION OF SUCCESS IN MILITARY JOBS

Torsten Husen
Institute of Educational Research
Teachers College, University of Stockholm
Sweden

Insufficient research has been done on the predictive validity of the interview for purposes of personnel selection. Few well-controlled studies have been carried out, and existing research reports often suffer from a failure to provide information on important relevant variables—what sort of interviewers were used, the type of procedure they followed, or the extent to which the results of standardized tests and questionnaires entered into the evaluations.

Validation studies of interview procedures should provide answers to the following two questions: (a) How effectively can an interviewer integrate qualitative and quantitative data to arrive at valid predictions of future behavior? (b) What is the predictive value of the interview as compared to other means of assessing personality and behavior? In setting up experiments to answer these questions, many variables must be taken into account. Among them are the age, sex, and training of the interviewers, the amount and type of information available from sources external to the interview itself, the kind of subjects studied, and the purpose for which the interview was conducted. In large-scale assessment programs in which many interviews evaluate comparable samples of subjects, all these factors, with the exception of the interviewer, can be held approximately constant. This permits, as in the studies described here, evaluation of the interviewer as an independent variable. To the author's knowledge, little research has hitherto been conducted on the extent to which interview results are affected by the personal characteristics of the interviewer.

This report deals with the findings of several research projects carried out for the Personnel Selection Department of the Swedish Armed Forces in 1952-1955 by S. E. Henricson and the author. The studies described were designed to throw light on the various questions cited above.

Study 1

Subjects

The subject source consisted of Swedish military conscripts called up in the year 1950—a group comprising about forty-two thousand young men 19 years of age. All those who were physically fit, and who had obtained IQ's of 100 or above on the group intelligence test given at the time of induction, were selected to undergo interviews designed to screen out individuals suitable for NCO and officer training. The reduced sample consisted of approximately twenty-one thousand men. Of them, 6,079 subjects were selected on the basis of the interview for primary NCO train-

ing, and 3,564 were selected for an advanced NCO training program which included basic training for officers.

About two-thirds of the "advanced" trainees had passed the Swedish Matriculation examination, or possessed equivalent schooling. The mean IQ of the sample was 115, and the S.D. was approximately one-third that of the total group of conscripts. Of the group of candidates, 3,003 started their basic training about seven months after the registration and selection procedure, the rest being called up one year later. Criterion data were therefore available for only 3,003 of the advanced trainees. These subjects were not, it should be noted, selected from the original group of candidates on the basis of estimated leadership capability.

Interviewers

Forty-nine individuals with varied educational and experiential backgrounds served as interviewers. Twenty-one of the group were regular service officers, twenty-one were male civilians, and seven were female civilians. Of the twenty-one male civilians, thirteen had had university training in psychology up to or beyong the B.A. degree, and the remaining eight had had experience in social work or similar fields. Four of the seven female interviewers had had university training, and three had had experience in social work. All interviewers were given a three-week course in the technique of interviewing shortly before starting their jobs.

Interviews

The interviews were about 15 minutes in length. Each interviewer evaluated between four and five hundred randomly assigned men. At the end of the interview, each subject was rated by the interviewer on a five-interval scale pertaining to leadership potential. The scale distribution was forced, the interviewers being requested to assign 5 per cent, 20 per cent, 50 per cent, 20 per cent, and 5 per cent of their interviewees, respectively, to the five ordered classes of the scale. Only those subjects assigned to the highest two categories were subsequently considered for training. On the average, each interviewer selected from his own group of conscripts about 100-125 men for primary, and about 50 men for advanced NCO training.

Criterion Data

The NCO training was carried out at 48 different centers. The candidates at each center were rated on achievement by their commanding officers on three different occasions: After about one-third of the training had been completed, after about two-thirds had been completed, and at the end of the course. These ratings were made in terms of a 10-point scale where scores of four or below represented unsatisfactory performance.

The criterion used for purposes of comparative evaluation of the interviewers was failure rate, i.e., the percentage of NCO candidates selected who subsequently received unsatisfactory ratings in training. Drop-outs for psychological reasons were included in this rate, but failures due to physical disease or disability were not.

Results

For the two groups of trainees studied, the overall failure rates were 7.7 per cent for the primary, and 5.6 per cent for the advanced candidates. There was

considerable variation in the rates among different training units, the S.D. of the advanced-group distribution, for example, being 6 per cent. This figure differs significantly from that which would be expected on the assumption of equal probability of failure in the various units. The standadrs of judgment used by the commanding officers can therefore be assumed to have differed from unit to unit. To cite an illustrative difference, the failure rate for the artillery was 11.2 per cent, as against 5.2 per cent for the infantry.

Comparison of the failure rates for the various interviewers revealed marked individual differences. The range was from 0 to 21 per cent, with an S.D. of 5 per cent. This variation also differed significantly from that expected on the basis of chance. Moreover, the difference remained significant when the "difficulty" of the various training units was held constant. (It should also be noted that if allowance was made for differences in skill among the various interviewers, there still remained a systematic variation in failure rate among the units.)

Table 1 presents end-of-training failure-rate figures for various categories of interviewers. Rates for both primary and advanced NCO trainees are given. As can be seen from the table, the differences in the success of the various groups of interviewers are small. Analysis of the achievement ratings made earlier in the

TABLE 1
PER CENT FAILURE IN NCO TRAINING FOR CANDIDATES SELECTED BY VARIOUS TYPES OF INTERVIEWERS

Type of interviewer	Primary NCO Training		Advanced NCO Training (including basic training for officers)	
	Number of interviewees selected	Per cent of trainees failed at end of 12 months	Number of interviewees selected	Per cent of trainees failed at end of 15 months
Regular officers	2614	7.5	1537	6.0
Male civilians with some post-graduate training	2160	7.4	775	5.2
Male civilians with some university training or experience in social work	450	8.7	172	5.8
Female civilians with university training	476	8.0	366	6.3
Female civilians with experience in social work	379	8.2	153	4.0
Total:	6079	7.7	3003	5.6

course of training showed larger differences between the groups. In interpreting the presented end-of-training differences, account should be taken of the fact that the differences between the training units with respect to standards of judgment tend to be more pronounced in the latter part of training. During the initial, basic-training period, the demands made on the candidates are more alike than during the final period, when emphasis is on instruction in the special duties of platoon commanders.

Analysis of failure-rate data for the basic-training period also showed the female interviewers to have the lowest and the regular officers to have the highest failure rates, with the civilian male interviewers occupying an intermediate position. This finding was somewhat unexpected, since it is natural to anticipate that officers, with their many years of military experience and their familiarity with training requirements, would make better predictions than civilians. Moreover, the officers used had been selected by their superiors as being especially suitable for the task of interviewing. It should not be forgotten, however, that all the interviewers had had lectures on the organization of training and the demands of various assignments during their preparatory course in interview techniques.

In summary, the findings of Study 1 suggest that interviewers with first-hand experience of the job for which selection is being carried out do not give significantly more valid ratings than individuals with basic psychological training, that persons with university training in psychology do not make better interviewers than those with experience in social work, and that female interviewers seem to succeed as well as male interviewers.

Study 2

The primary aim of investigation was to determine how the predictive value of interview ratings compares with that of standardized psychological tests.

Subjects and Procedure

The selection of officer candidates was the concern of the study. The subject sample was comparable to the NCO group of the first investigation with respect to social background, amount of education, and distribution of measured intelligence.

As part of the selection process, all subjects were administered a battery of objective tests comprising one conventional group intelligence test, the Bennett Test of Mechanical Comprehsneion, a schedule dealing with social attitudes, and a personality inventory measuring introversion-extroversion and neuroticism. Each subject also filled out a 55-item biographical questionnaire dealing with social background, education, and previous vocational experience, and wrote a brief autobiography.

The task of the interviewer was to integrate the candidates' test results with impressions obtained in conversation to form a global estimate expressible in the form of a rating. Each interviewer saw about five to six candidates per day, and had about 60 minutes at his disposal for each interview. Before starting a session, the interviewer spent about 30 minutes studying the subject's test results and autobiography. After the interview, he devoted about 25 to 30 minutes to writing a short personality description according to a standard format, and rating the interviewee with respect to estimated suitability as an officer, and ability to profit from academic instruction. These ratings were made on a nine-point scale according to a forced-normal distribution.

Two criteria of success in training were used: (a) normalized rank order in officer candidate school at the end of the six-month basic training period, and (b) average of the grades received in the two most "theoretical" courses ("Tactics," and "Technology of Arms").

The analysis of the data consisted of computing, for the subjects seen by each interviewer and for the subject sample as a whole, correlations between: (a) rated academic aptitude and average of the candidates' "Tactics" and "Technology of Arms" course grades, (b) ratings of suitability as an officer, and rank order in officer candidate school, (c) average of standard scores received on the group intelligence and Mechanical Comprehension tests, and average of grades in the two courses mentioned above, and (d) average of intelligence, Mechanical Comprehension test standard scores, and rank order in officer candidate school. The difference between correlations (a) and (c), and between correlations (b) and (d) were also determined.

Results

Results for the 22 interviewers who had assessed a total of 50 or more candidates are given in Table 2. Fisher's z-transformation was utilized in computing the correlations therein presented.

The findings of the study are summarized as follows:

1. Comparison of the total-group validities for the course-grade criterion shows that, taken together, the two ability tests predict grades in the courses considered, as well as do interview ratings based on subjective impressions plus consideration of all of the test scores, inventory results, and autobiographical data which were available for the subjects.

2. The overall validity of the interview ratings for predicting rank order in officer candidate school at the end of six months of training is significantly higher (.01 level of confidence) than that obtained for the ability tests. Moreover, study of the correlations for the various subgroups of subjects rated by the individual interviewers shows that the interviewers tended to provide consistently better predictions than the tests.

In connection with these findings, it should be mentioned that for a second group of subjects rated by 16 less experienced interviewers, the validities for the same criterion were .44, and .44 for the interview ratings and ability-test averages, respectively. In view of the consistency of the present results, it seems reasonable to suppose that differences in the amount of experience on the part of the two groups of interviewers were responsible for the observed differences in the predictive value of their ratings.

It has often been questioned whether an interviewer is capable of making predictions better than those obtainable from objective test scores. The findings of this study appear to indicate that the answer to this question depends on a number of factors, among which are the nature of the criterion and the skill of the interviewer. The results showed that when the criterion consisted of course grades test scores yielded predictions as valid as those made by the interviewers, even though the interviewers' predictions were based on a considerably larger amount of data. When the more complex criterion of rank order in officer candidate school was utilized as a measure of success, however, the interviewers' predictions were

TABLE 2

COMPARATIVE VALUE OF INTERVIEWERS' RATINGS AND OBJECTIVE TEST SCORES FOR PREDICTING SELECTED COURSE GRADES AND RANK ORDER OF OVERALL PERFORMANCE IN OFFICER CANDIDATE SCHOOL

		Correlations between:					
viewer	Number of S's	Rated academic aptitude and average of 2 course grades (1)	Rating of suitability as an officer and rank order at end of basic training (2)	Average of 2 test scores and average of 2 course grades (3)	Average of 2 test scores and rank order at end of basic training (4)	Diff: Cols. (1-3) (5)	Diff: Cols. (2-4) (6)
1	80	.34	.45	.26	.18	+.08	+.27
2	93	.60	.49	.59	.41	+.01	+.08
3	50	.38	.51	.44	.25	-.06	+.26
4	89	.55	.52	.56	.45	-.01	+.07
5	132	.46	.54	.46	.42	.00	+.12
6	118	.52	.56	.59	.48	-.07	+.08
7	169	.51	.42	.53	.48	-.02	-.06
8	135	.71	.61	.61	.42	+.10	+.19
9	152	.53	.42	.52	.48	+.01	-.06
10	116	.44	.41	.50	.41	-.06	.00
11	193	.42	.40	.41	.31	+.01	+.09
12	68	.50	.60	.34	.20	+.16	+.40
13	59	.41	.56	.36	.26	+.05	+.30
14	117	.53	.41	.42	.41	+.11	.00
15	115	.57	.51	.58	.44	-.01	+.07
16	61	.67	.58	.59	.58	+.08	.00
17	99	.40	.47	.49	.35	-.09	+.12
18	111	.46	.48	.52	.45	-.06	+.03
19	69	.44	.54	.54	.38	-.10	+.16
20	119	.46	.48	.34	.27	+.12	+.21
21	132	.56	.49	.50	.44	+.06	+.05
22	158	.59	.44	.35	.41	+.25	+.09
Total:	2435						
Values for engire group:		.52	.49	.50	.41	+.02	+.08

found to be significantly (and consistently) more valid than those derived from objective test scores. Comparative data suggested, in addition, that differences in the amount of experience possessed by interviewers may affect the validity of their ratings.

Study 3

In order to throw light on the contribution of individual interviews to the prediction of success in pilot training, another analysis was undertaken. The study was devoted to determining the influence of interviewers' evaluations on prediction by the Swedish Air Force Selection Board of a pass-fail criterion of success in pilot school.

The pilot selection procedure is a two-day process comprising physical tests, psychological tests, and interviews. During the first day, the candidate takes an extensive battery of group tests and undergoes a medical examination. The following day, he is interviewed for about one hour by a trained interviewer, and then appears before the entire Selection Board for a short group interview.

Before seeing each candidate, the Selection Board is provided with the personality description and the ratings (on a nine-point scale) made on the basis of the individual interview. After its own interview, the Board gives the candidate an overall rating consisting of the average of the individual ratings made (also on a nine-point scale) by its various members.

It can be seen that in this procedure the interviewers can influence the Board, but the Board does not influence the ratings of the interviewers. It is therefore possible to partial out the effects of interviewers' evaluations so as to determine the amount of their contribution to the final assessments made by the Selection Board.

Tables 3 and 4 present biserial correlations against the pass-fail criterion of: (a) various interviewers' ratings, and (b) ratings made by the Selection Board before and after partialling out the variance attributable to the evaluations of the interviewers. Both sets of data, which are based on different samples, show that the predictive efficiency of the Board ratings tends to vary with the validity of the interviewers' assessments, and that the validity of the Board assessments is reduced by eliminating the influence of the interviewer. It appears, then, that the individual interviews do contribute to the efficiency of the Board predictions.

TABLE 3
BISERIAL CORRELATIONS OF INTERVIEWER AND SELECTION BOARD RATINGS WITH PASS-FAIL CRITERION OF SUCCESS IN PILOT SCHOOL

Interviewer	Number of pilots	r_{bis} of interviewer ratings	r_{bis} of Selection Board ratings
A	69	.21	.38
B	27	.00	.05
C	69	.33	.40
D	27	.50	.65
E	39	.36	.27
F	61	.49	.45

594869 O - 61 - 12

TABLE 4

BISERIAL CORRELATIONS SHOWING COMPARATIVE VALUE OF INTERVIEWERS' RATINGS, AND OF SELECTION BOARD RATINGS WITH AND WITHOUT INTERVIEWERS' RATINGS, FOR PREDICTING PASS/FAIL CRITERION OF SUCCESS IN PILOT SCHOOL

		r_{bis} of Selection Board ratings with:	
Interviewer	r_{bis} of interviewer ratings	Interviewers' ratings considered	Inverviewers' ratings partialled out
1	.62	.69	.39
2	.46	.51	.27
3	.40	.46	.25
4	.39	.34	.11
5	.12	.12	.07

Discussion

On the whole, the results of these investigations cannot be interpreted as completely discouraging to the advocate of the interview as a personnel selection method. In concluding this communication, two additional points should be made. The first relates to the effects of interviews on the motivation of the applicants. In yet another study, two equivalent groups of subjects were selected, purely on the basis of their objective test scores. One group was then given a 15-minute interview during which they were encouraged to express their views on their assignments. The second group received no interviews, and selection was in no case influenced by the interview results. Several months later, a survey revealed 20 per cent more satisfaction with their assignments among the subjects who had been interviewed as compared with those who had not. This would appear to indicate that interviews can make an effective contribution to morale.

A second, related point is that interviewers frequently have on hand biographical information relating to the suitability of the subject for the job to which he has been assigned. By taking a "therapeutic" approach, the interviewers can, using these data, attempt to influence the individual's attitude toward his assignment.

Because of these and other advantages to be derived from the interview, it is the author's belief that the procedure can contribute importantly to military personnel programs. Research efforts should be concentrated on determining what sort of technique, and what sort of data obtainable by that technique, will yield the most valid interview estimates.

Summary

The paper reports the results of three studies on the value of the interview for purposes of military personnel selection. Attention is focused on three main questions: (a) How valid are estimates based on interviews for predicting success in various types of military training? (b) How does the interview compare as to validity with other methods of predicting training success? (c) Does the validity of interview ratings vary with the personal characteristics of the interviewer?

Study 1 (concerned primarily with Question c) dealt with NCO trainees. From among twenty-one thousand Swedish military conscripts called up in 1950 who were physically fit and had IQ's of 100 or above, 6,079 men were selected for primary and 3,564 were selected for advanced training as non-commissioned officers. Criterion data were available for all of the former, and for 3,003 of the latter group.

The selection was done on the basis of 15-minute, individual interviews carried out by 49 interviewers of varied educational and experential backgrounds. Twenty-one of the interviewers were military officers, twenty-one were male civilians with university training in psychology or experience in social work, and seven were female civilians with university training or social work experience. Each interviewer saw a total of four to five hundred randomly assigned men, whom he rated on a five-interval scale of leadership potential according to a forced-normal distribution. Those subjects falling into the highest two intervals were subsequently selected for training.

The criterion data consisted of achievement ratings made on a ten-point scale, during and at the end of training, by commanding officers at the 48 training centers to which the candidates were assigned. Failure rate (the percentage of trainees who: (a) received criterion ratings of four or below, or (b) dropped out for other than physical reasons) constituted the measure in terms of which the performance of the interviewers was compared.

Analysis of the end-of-training data revealed only small differences among the various categories of interviewers with respect to the percentage of primary and advanced NCO candidates selected who subsequently failed in training (Table 1). Differences in predictive accuracy among individual interviewers were nevertheless marked, and significantly exceeded chance expectation, even when allowance was made for differences in standards of evaluation among the various training centers.

Analysis of achievement ratings made during the course of training revealed larger differences in failure rate among the various categories of interviewers. It also showed that the female interviewers made more accurate predictions than the male civilians, who in turn performed better than the officers.

Increased differences in judgmental standards between centers during the latter, more specialized part of the training are posited in partial explanation of the smaller differences among interviewer groups found for the end-of-training data. In general, the results of the study showed that the predictive value of interview ratings does not markedly vary with such interviewer characteristics as sex membership, training in psychology and social work, or first-hand experience with the assignments for which selection is being carried out. In connection with the last point, attention is drawn to the unexpectedness of the finding that experienced military officers familiar with NCO training requirements did not make better predictions than others.

The purpose of Study 2 was to determine how the predictive value of interview ratings compares with that of objective psychological tests. The subject sample consisted of 2,435 officer candidates. Data available for these subjects comprised: (a) scores on various tests (indlucing a group intelligence test and the Bennett Test of Mechanical Comprehension), (b) responses to a 55-item biographical questionnaire, (c) short autobiographical essays, (d) ratings of "suitability as an officer" and of "ability to profit from academic training" made on a nine-point scale after a one-hour interview and study of the candidate's test scores and biographical data, (e) normalized rank order in officer candidate school at the end of the six-month basic training period, and (f) average of the grades received in two theoretical courses ("Tactics" and "Technology of Arms").

The analysis of the data consisted of computing (for the candidates seen by each interviewer, and for the subject sample as a whole) correlations against the course-grade and rank order criteria of: (a) the two types of interview ratings, and (b) the average of the subjects' standard scores on the two ability tests specified above. The differences between the correlations of the ratings and of the test-score averages against each criterion were also calculated (Table 2).

Comparison of the total-group correlations for the course-grade criterion revealed no significant difference in predictive accuracy between the ability test scores and the interview ratings of academic aptitude. Rank order in officer candidate school, however, was predicted significantly less well (.01 level of confidence) by the ability test scores than by the interview ratings of suitability as an officer. This difference dropped to zero in a comparative study in which 16 less-experienced interviewers were employed. Taken together, the findings suggest that the relative efficiency of objective test scores and interview ratings depends on several factors, among which are the nature of the criterion and the skill of the interviewer.

Study 3 was designed to determine the degree to which the results of individual interviews contribute to the validity for predicting success in pilot training of assessments made by the Swedish Air Force Selection Board. A candidate passing theough the Swedish pilot selection procedure must appear before the Selection Board for a brief group interview, after having undergone a series of physical and psychological tests and an individual interview of one hour's duration. Before seeing each candidate, the Board is provided with the ratings and personality description made by the individual interviewer, whose effect on the Board ratings can accordingly be determined by partial correlation procedures.

Tables 3 and 4 present biserial correlations against a pass-fail criterion of success in pilot school for: (a) individual interviewers' ratings, and (b) Selection Board ratings before and after partialling out the variance attributable to the interviewer. The data show that the predictive efficiency of the Board assessments varies with the validity of the interviewers' ratings, and is substantially reduced when the influence of the interviewer is eliminated. It can be concluded that individual interviews do contribute to the efficiency of the Board predictions.

In conclusion, the author points out that concern with the predictive aspects of the interview should not be carried to the point of obscuring its other potential contributions to military personnel programs. Interviews can be used, among other things, to improve morale and to manipulate attitudes and motivation. It is suggested that future research be focused on improving interview techniques and determining the most useful kinds of information obtainable through their use.

Résumé

Le rapport présente les résultats de trois études portant sur la valeur de l'interview pour la sélection de personnel militaire. L'attention est centrée sur trois questions principales: (a) Quelle est la valeur des évaluations basées sur des interviews pour prédire le succès dans divers types de formation militaire? (b) Quelle est la valeur de l'interview par rapport à celle d'autres méthodes de prédiction? (c) Est-ce que les estimations déduites des interviews varient en valeur selon le type d'examinateur?

L'étude 1 (qui traite surtout de la question n° c) concerne des candidats sous-officiers. Sur environ vingt et un mille recrues suédoises, appelées sous les armes en 1950, qui étaient physiquement aptes et avient des quotients d'intelligence de 100 ou plus, 6.079 ont été sélectionnées pour la formation élémentaire et 3.564 pour la formation plus poussée de sous-officiers. Il existait des résultats du succès dans la formation pour tous ceux de premier groupe et pour 3.003 candidats du second.

La sélection fut faite sur la base d'entrevues individuelles de 15 minutes effectuées par 49 examinateurs de formations diverses. Vingt et un des examinateurs étaient des officiers, vingt et un étaient des civils ayant une formation universitaire en psychologie ou de l'expérience dans le domaine des oeuvres sociales, et sept étaient des éléments féminins de niveau universitaire ou ayant de l'expérience dans le domaine du travail social. Chaque examinateur a interrogé au total quatre à cinq cents hommes affectés au hasard, qu'il a cotés suivant leurs capacités de commandement sur une échelle à cinq catégories, selon une répartition normale forcée. Les sujets tombant dans les deux catégories supérieures furent sélectionnés ensuite pour la formation.

Le critère du succès consistait en cotes effectuées sur une échelle de dix points, pendant et à la fin de la période de formation, par des officiers commandants des 48 centres d'entraînement auxquels ont été affectés les candidats. Le taux d'échecs (le pourcentage de candidats qui (a) ont obtenu quatre points ou moins, (b) ont échoué pour d'autres motifs que leur état physique) constituait la mesure selon laquelle on comparait les jugements des examinateurs.

L'analyse des données pour les deux groupes de candidats n'a révélé que de petites différences entre les diverses catégories d'examinateurs en ce qui concerne les taux d'échecs à la fin de la formation (Table 1). Des différences entre les examinateurs individuels ont été néanmoins marquées, et sont restées statistiquement significatives même après avoir tenu compte des différences dans les normes d'évaluation entre les divers centres d'entraînement.

L'analyse des cotes de résultat faites pendant la durée de la formation a révélé de plus grandes différences dans le pourcentage d'échecs entre les diverses catégories d'examinateurs. Elle a également montré que les examinatrices faisaient des jugements plus exacts que les examinateurs civils qui, à leur tour, ont donné de meilleures estimations que les officiers.

De plus grandes différences entre les centres d'entraînement en ce qui concerne les normes d'évaluation de la réussite pendant la dernière partie, plus spécialisée, de la formation, sont avancées comme explication partielle des plus petites différences entre les groupes d'examinateurs constatées pour les données de

fin de formation. En général, les résultats de l'étude ont indiqué que la valeur des cotes d'entrevue ne varie sensiblement pas en fonction des caractéristiques de l'examinateur telles que le sexe, les connaissances en psychologie et le travail social, ou l'expérience pratique dans le domaine pour lequel la sélection est faite. Ici l'on note le caractère inattendu de la constatation que les officiers militaires, quoique bien familiarisés avec les exigences de la formation de sous-officiers, n'ont pas fait de meilleures prévisions que les autres.

Le but de l'étude n° 2 consistait à déterminer quelle est la valeur de prévision des interviews par rapport à celle de tests psychologiques objectifs. Le groupe de sujets se composait de 2.435 candidats officiers. Les données dont on disposait pour ces candidats comprenaient: (a) les résultats de divers tests (y compris un test d'intelligence de groupe et le test de compréhension mecanique de Bennett); (b) les réponses à un questionnaire biographique de 55 items; (c) de brèves autobiographies écrites par les sujets; (d) des cotes d'"aptitude pour être officier" et d'"aptitude académique" établies par les examinateurs sur une échelle de neuf points après des entrevues d'une heure et étude des résultats des tests et des données biographiques; (e) la classification à l'école des candidats officiers à la fin de la période d'instruction de base; et (f) la moyenne des résultats obtenus à deux cours théoriques ("Tactique" et "Technologie des Armes").

L'analyse des données a consisté à calculer (pour les candidats interviewés par chaque examinateur et pour le groupe de sujets tout entier) les corrélations avec les critères de résultats des cours et de la classification à l'école (a) des deux genres de cotes faites par les examinateurs, et (b) de la moyenne des notes standardisées des sujets pour les deux tests de capacité susmentionnés. Les différences entre les corrélations des cotes et des tests avec chaque critère ont également été calculées (Table 2).

La comparaison des corrélations basées sur le groupe total pour le critère de résultat des cours n'a pas revile de différences significatives de valeur de prévision entre les tests de capacité et les cotes d'interview quant à l'aptitude académique. La classification à l'école des candidats officiers, cependant, a été prédite moins bien (0,01 niveau de confiance) par les tests de capacité que par les cotes d'aptitude pour être officier. Cette différence est tombée à zéro dans une étude comparative basée sur 16 examinateurs moins expérimentés. Prises dans leur ensemble, les constatations montrent que la valeur relative des résultats objectifs des tests et des cotes d'interview dépend de divers facteurs, entre lesquels sont la nature du critère et la compétence de l'examinateur.

L'étude n° 3 avait pour but de déterminer la mesure dans laquelle les entrevues personnelles contribuent à la valeur des estimations faites par la Commission de Sélection de l'Armée de l'Air suédoise. Un candidat subissant la sélection suédoise des pilotes doit comparaître devant la Commission de Sélection pour répondre à quelques questions après avoir passé par une série de tests physiques et psychologiques et une entrevue individuelle d'une durée d'une heure. Avant d'examiner chaque candidat, la Commission reçoit les cotes et la description de l'individu, établies par chaque examinateur, dont l'effet sur les évaluations de la Commission peut en conséquence etre determiné par la technique de corrélation partielle.

Les Tables 3 et 4 présentent des corrélations bisérielles pour le critère de succès/échec à l'école des pilotes, (a) des cotes attribuées par des examinateurs individuels et (b) des cotes attribuées par la Commission de Sélection avant et après

déduction de la variance attribuable à l'examinateur. Les données montrent que l'efficacité de prévision des estimations de la Commission varie suivant la valeur des cotes des examinateurs, et est fortement réduite quand l'influence de l'examinateur est supprimée. On peur conclure que des entrevues individuelles améliorent les évaluations faites par la Commission.

Pour conclure, l'auteur souligne l'intérêt de ne pas insister sur le côté prévision de l'interview au point d'obscurcir ses autres contributions possibles aux programmes militaires de personnel. Les interviews peuvent servir, par exemple, à améliorer le moral et à analyser les attitudes et les motifs. Il est suggéré de faire porter les recherches futures sur l'amélioration de la technique de l'entrevue, et de déterminer les informations les plus utiles qu'elle est susceptible de fournir.

Commentary by Discussant

Dr. Jean Ungricht (Switzerland):

The problems attacked by Dr. Husen are highly important ones, and I think you will agree that studies of this kind are greatly needed. Despite the difficulty of empirically demonstrating its validity, the interview continues to hold a favored place in personnel selection procedures. Among the reasons for this, as Dr. Husen pointed out, are its value as a method of counselling and guidance, of communicating information, and of dealing with certain problems of attitudes and motivation. In addition, it provides a convenient means of assessing social acceptability and personal appearance and mannerisms—characteristics which may be important in jobs which require dealing with the public or fitting into pre-existing work teams. Because of its usefulness for these purposes, the interview probably never will be completely abandoned.

It must be admitted that face-to-face contact is an excellent means of evaluating certain types of overt behavior and surface personality characteristics. Their success at such tasks may explain why interviewers tend to become so strongly convinced of their insight and omniscience, and why there is a tendency to apply the interview to problems for which its value is unsubstantiated. One such problem is that of assessing specific abilities, a purpose for which the interview is used almost as often as it is for the more appropriate ends just mentioned. But, in this area, as innumerable validation studies have shown, the value of the interview as a predictive instrument is extremely questionable. If we analyze the processes involved, it is easy to see why this should be so. Halo and other rating errors, bias on the part of the interviewer, the difficulty of obtaining full and accurate information on relevant points—these and many other problems, so often described as to need no repetition here, tend to attenuate the validity of the interview for predicting success in many kinds of jobs. We must conclude, as Eysenck puts it in "Uses and Abuses of Psychology," that in this context the interview "is an almost complete failure, and that the time devoted to it, so far as accuracy of prediction goes, is wasted. The interview, far from increasing accuracy of prediction, may actually lesses it." Bellows and Estes come to a similar conclusion in their book, "Employment Psychology." Rather than recommending that the method be abandoned,

however, these authors "would prefer to encourage development of research on ways of conducting and evaluating the interview with emphasis on the use of more valid supporting data." I personally believe that this is the proper attitude to take, and think that the studies of Dr. Husen constitute a prototype for research and an important contribution to the existing body of empirical data.

In conclusion, I should like to put forward the following question for discussion. In research of the type just reported, is it not possible that we are tending to confuse the interview as a tool with the interviewer as a tool? Are we not actually concentrating on evaluation of the man rather than on evaluation of the technique? It cannot be overlooked in the present studies that some interviewers made much better predictions than others. This suggests that quite different results might be obtained if research comparing the relative predictive accuracy of tests and interviews were to begin by screening out the best interviewers from a large sample. Why should we not apply the same methods to the selection of interviewers as we do to the selection of items to be included in a test or tests to be included in a battery? Application of the traditional psychometric approach in research on the interview might lead us into new fields.

General Discussion

Dr. Robert Glaser (U. S. A.):

In a study recently completed by the American Institute of Research, results similar to those of Dr. Husen were obtained. We were concerned in this study with the selection of civilian supervisors at military depots. The first step was to administer various paper-and-pencil tests, and to obtain measures of on-the-job performance for a large sample of foremen. From this sample we selected two groups which were equivalent with respect to test scores, but which differed significantly on the criterion. The members of these groups were then given both individual interviews and group interviews carried out by three interviewers working jointly. In addition, they participated in group discussion sessions. Analysis of the resulting data permitted us to determine whether the successful foremen, who could not be differentiated from their less successful colleagues on the basis of the tests, could be distinguished by means of interviews or observation of group performance.

The finginds, to put it briefly, were negative. Very little was added to prediction by the interview or group discussion procedures. An interesting side observation was that the individual and group interviews yielded very similar results—a finding which suggests (for those who wish to continue using interviews) that the less costly, man-to-man procedure can safely be continued. It should also be noted that the group discussion procedure, in which one man observed the performance of four subjects, yielded somewhat better results than either type of interview. This method involves the lowest subject-to-examiner ratio of all, and, thus, is the least expensive to apply.

Mr. V. Coucheron Jarl (Norway):

We have also carried out, and with rather encouraging results, some research on the selection of NCO candidates using interviews and objective test vari-

ables. In previous factor-analytic studies, we had found that our usual ability tests did not cover that part of the criterion variance which we called "the sergeant factor." Hence, the interviews were specifically arranged to gather and systematize information on elements of this factor. The data thus obtained were put into the mill along with the test results and analyzed in exactly the same way. Factor analysis and multiple regression procedures both selected two interview variables as contributing significantly to prediction. One pertained to leadership and the other to social adjustment, as judged by the interviewer. The results of the study have not yet been cross-validated, but the findings for the experimental sample are hopeful. Personally, therefore, I am inclined to agree with Dr. Ungricht: emphasis should be placed on specifying the type of material to be collected in interviews, and on subjecting this material to the same treatment as that accorded to data obtained by other means. We should attempt to get away from the approach which extols global, intuitive integration of all kinds of things without knowing what it is that is integrated, or how the integration is carried out.

Dr. H-K Knoepfel (Switzerland):

I should like to stress the point that the interview involves interpersonal relationships which vary according to the personality of both the interviewer and the interviewee. Interviews can yield assessments of social behavior, but this can be done reliably only if the interviewer has adequate insight into his own personality and into how this behavior influences that of the interviewee. Psychiatric training can be of value in achieving such insight.

Another point is that interviews should not be cut too short. A rushed interview can degenerate into a series of yes-no responses which yield little information on the personal and social characteristics about which the interviewer is interested in learning. We have followed the principle of interviewing only one individual a day, and have had good experience with interviews three or four hours in length. Of course, this is possible only if the number of candidates to be examined is very small.

Reply

Dr. Torten Husen:

I wish to thank those who have commented on my paper and to express my general agreement with the remarks of the discussant. I did not come here to kill the use of the interview in personnel selection, and personally think that it can have much to contribute.

Dr. Ungricht has suggested the possibility of selecting outstandingly good interviewers for purposes of study. Unfortunately, it is not always easy to do this. The main problem is that interviewer excellence tends to vary with features specific to the interview situation and to the criterion to be predicted. Some interviewers tend to be better for some purposes, and some for others.

In my opinion, it is of primary importance to conduct research on the nature of the processes which occur during the interview itself. We know a good deal about the predictive value of interview assessments, but few studies have hitherto been done on such problems as to how the impressions built up during an interview are derived. Interviewing is a technique with values other than the purely predictive, and it is important to know more about what actually happens in interpersonal situations of this kind.

SESSION IV

Chairman: Dr. Norman A. B. Wilson

Senior Psychologist, The Admiralty

London, England

SOME HUMAN FACTORS IN NOISE CONTROL

K. D. Kryter
Operational Applications Laboratory
Air Force Cambridge Research Center
Bolling Air Force Base, Washington, D. C.

Noise is almost universally present in military operations, and it is mandatory that communidation systems be developed which will be reliable and accurate in such noise. In order to assist the engineer to this end, and to extablish criteria of satisfactory performance, attempts have been made to use acoustical measures of background noise and the speech signal to predict the intelligibility or understandability of speech. In addition, measurements have been made of the relation between the noise spectrum and subjective feelings of annoyance due to noise. It is the purpose of this paper to describe how the results of such experiments are applied by human engineers to the solution of problems arising from the presence of noise in the environment of the human operator.

Communication in Noise

Primarily on the basis of research done at the Bell Telephone Laboratories (French & Steinberg, 1947), it has been possible to divide the frequency spectrum of speech sounds into 20 bands, each of which is equally important to the intelligibility of speech. In a given situation, the contribution of each band to intelligibility is a function of the ratio (expressed in decibels) between the speech signal and the background noise simultaneously reaching the ears of the listener. Perfect reception, arbitrarily represented by a value of 1.0, is achieved for any one band at a signal-to-noise ratio of + 18 db; zero intelligibility for that band is present when the signal-to-noise ratio is -12 db. The percentage contribution for each of the 20 bands is a linear function between these two extreme signal-to-noise ration. Hence, by measuring the speech and noise spectra reaching the listener's ears, and making certain simple computations within each frequency band, one can arrive at an "Articulation Index" of the understandability of speech for any given communication system.

It will be recognized that this method assumes that the masking of a portion of the speech spectrum by noise has the same effect on intelligibility as the removal of that portion by frequency distortion or by filtering of the speech signal. This is a fairly sweeping assumption which in recent years has been subjected to intensive test. In general, it has been found that the assumption is justified when the noise spectrum is continuous. It does not hold, however, when only very narrow bands of noise are present. Narrow-band noise can result in a spread of masking high frequencies by low frequencies, and/or in what has been called a "remote masking" of low frequencies by high, and these effects when present must be taken into account. Thus, a recent experiment carried out in the Operational Applications Laboratory showed that if the 20-band method is to predict accurately the effects of low frequency noise on high frequency speech, and vice versa, the calcu-

lations must be based on the masking spectrum of the noise, rather than the acoustical spectrum. The masking spectrum, incidentally, was estimated in this study from previously reported data obtained with the use of pure tones, and the findings accordingly provide an interesting example of how the results of pure-tone research can be generalized to complex perceptual problems involving speech.

It should also be noted here that the Index method best predicts understandability for samples of speech sounds approximating the complexity of conversational speech. The intelligibility of various consonants has been shown to be differently affected by the same type of noise, and it has been recently found that vowel sounds are affected differently than consonants.

Figure 1 shows the relationship between various values of the Articulation Index and the percentage of different types of messages correctly understood under laboratory testing conditions. It can be seen from this figure that the type of speech material transmitted has an important effect on understandability. It has become fairly common engineering practice to expect a communication system having an Index of .4 to provide satisfactory intelligibility for connected discourse, or normal telephonic speech.

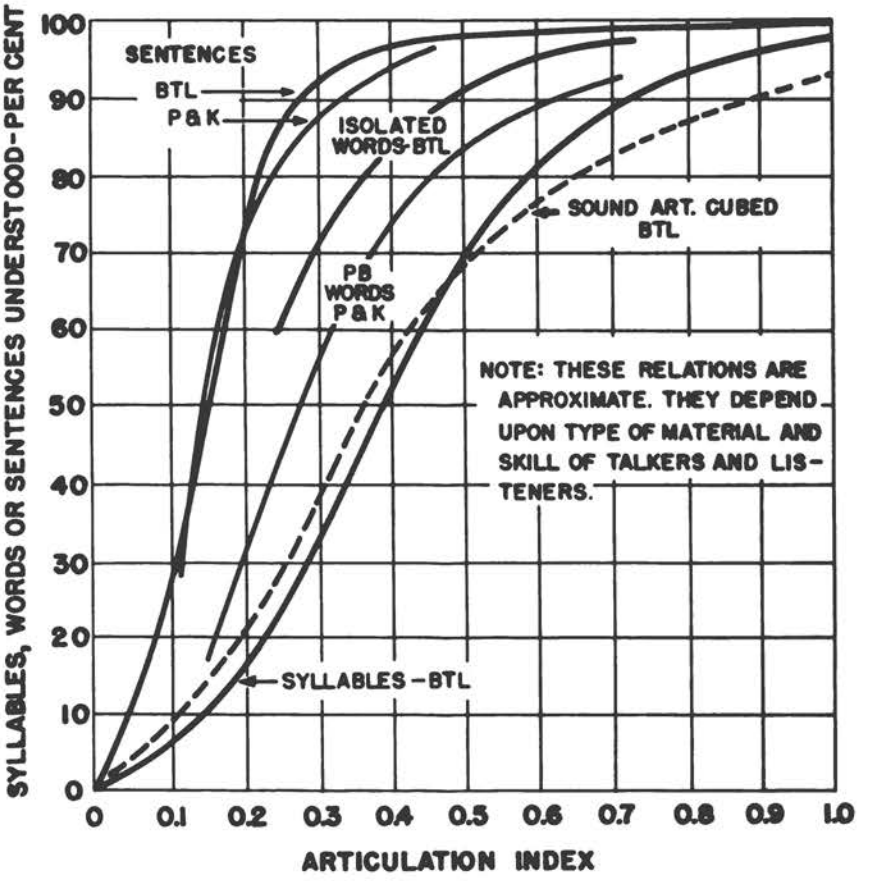


Fig. 1.--Relation between Articulation Index and Percentage of Test Material Correctly Perceived

In summary, it can be said that under the conditions of noise and frequency distortion usually present in military communication systems, the above method of calculating speech intelligibility indices from acoustical measures is sufficiently accorate for practical design purposes (Beranek, 1947). Figure 2 illustrates how the technique can be applied, for example, to aircraft inerphones. In the work sheet shown in this figure, the frequency range of audible stimuli (divided into 20 bands) is represented on the abscissa, and stimulus intensity in decibels is represented on the ordinate. The Articulation Index is obtained by taking the average ratio, over the 20 bands, of the speech and noise spectra represented by the two curves in the body of the graph.

It would appear at this stage that the job of human engineering is relatively straightforward. All that is needed is to make some simple acoustical measurements of the noise and speech signals reaching the ear of the listener. Then, if the Index computed from these measures is too low, the only need is to specify what reductions in the noise spectrum or what increases in the speech spectrum will be required to achieve a satisfactory level of speech intelligibility. In many cases, however, reducing the noise level or increasing the speech level may be impractical, or turn out to be too costly. Furthermore, it is sometimes impossible to make the speech signal more intense than the noise without running the risk of damaging the listener's ear with the speech signal. This is the case in certain environments aboard warships or near military aircraft, where noise levels can reach intensities of 130 db or more. Unfortunately, the performance of military operations under stressful conditions such as these often depends on the successful communication of speech.

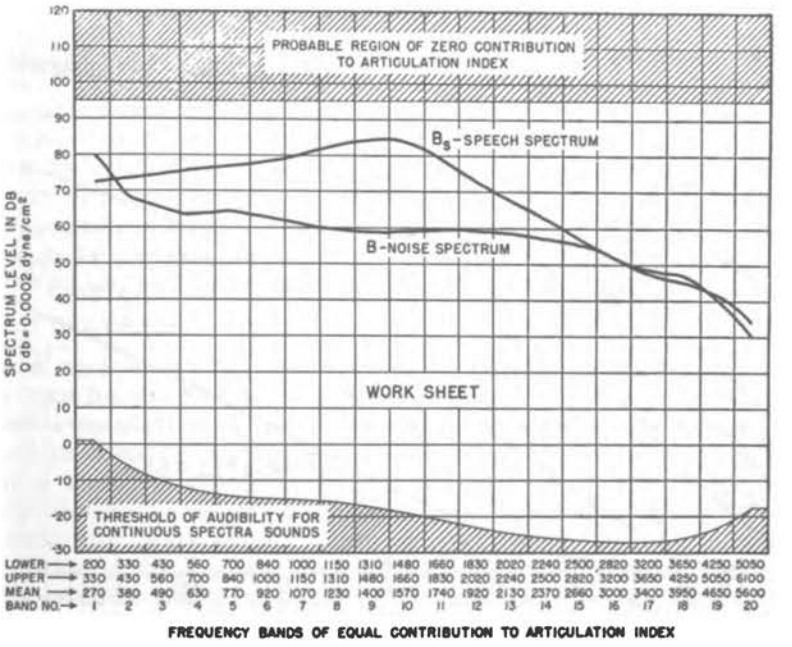


Fig. 2.--Work Sheet for Evaluating the Articulation Index of a Communication System under Noise.
B_s = long-term rms sound pressure level of speech plus 12 db. From Beranek (1947).

If communication requirements cannot be satisfied by making physical changes in the system, it is usually possible to improve matters by designing the language or messages to match man's perceptual capabilities as effectively as possible. For example, it is characteristic of human communication that the greater the redundancy in a message, the easier it is to communicate that message. This was illustrated in Fig. 1, which showed that sentences (or even long words) are more resistant to masking by noise than are individual words or syllables. In the terms of information theory: as the redundancy in a message is increased, the information content of the individual words or speech sounds is decreased. As a result, the meaning of an entire message can often be understood when only a few of the words are received.

Reducing information content by adding redundancy in the form, or words, or sentence structure does increase the intelligibility of speech messages in noise, but this process is usually expensive in terms of the time required to communicate, and, perhaps, in terms of the time required for the training of talkers and listeners—two factors which are often important in military operations. Another approach to the problem is to engineer the language so as to fix the size of the vocabulary or "message set" at the minimum required to meet operational needs for information. Here, a reduction in information content is effected by decreasing the number of words to be used, but the effect on speech intelligibility is the same. This is shown graphically in Fig. 3, which was taken from a study by Miller, Heise, and Lichten (1951). It will be seen that the intelligibility of the same words heard in noise was radically changed when the listeners knew the message set being used by the talker. The smaller the set, the higher the intelligibility for a given signal-to-noise ratio.

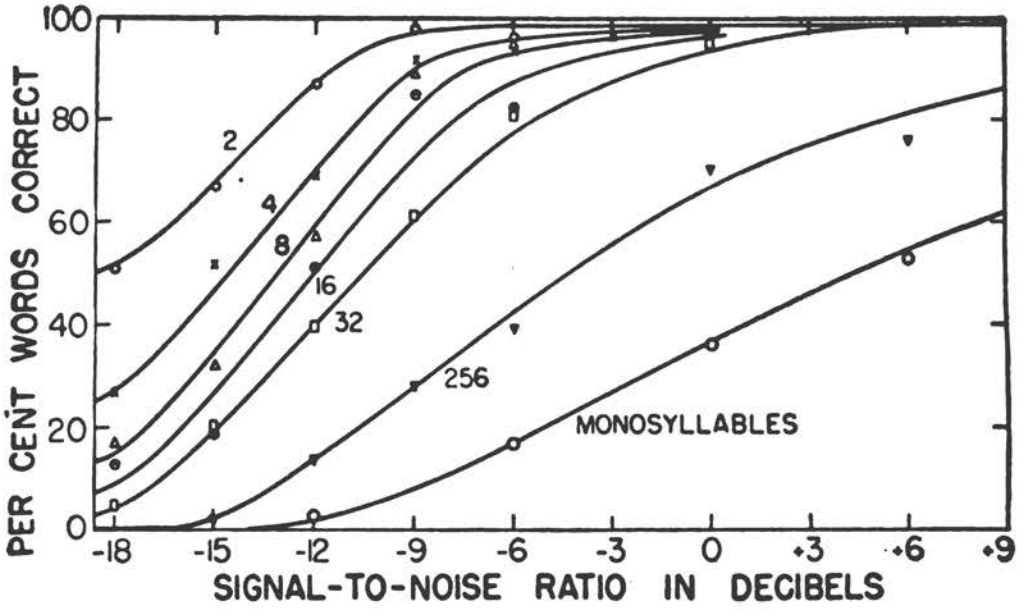


Fig. 3.--Word Intelligibility Scores Obtained at Various Speech-to-Noise Ratios for Test Vocabularies Containing Different Numbers of Alternative English Monosyllables.
The bottom curve was obtained with a vocabulary of approximately 1,000 monosyllables. From Miller, Heise, and Lichten (1951).

The USAF recently put the results of this study to good use in the operation of its communication network used in Air Defense. It was found that, because of the relatively high noise levels and poor frequency response present in the telephone lines used for "telling" radar information among aircraft control stations, the communication Index (computed according to the method previously described) was less than .4, and, thus, presumably unsatisfactory. Exorbitantly expensive steps were proposed to increase the power levels in the system in order to achieve a greater signal-to-noise ratio, and thereby increase the Index to at least .4. However, an analysis by military psychologists of the information to be transmitted over the system revealed that a good job of language engineering would be of considerable help. As a result, the size of the message set prescribed for use over these lines was reduced to about 32 different words—the digits, direction or heading, plus a few code words. This would be about "5 bits" worth of information in terms of information-theory measures (where bit = $\log_2$ of N alternatives). Restricting the messages to this small number of words permitted satisfactory operations, with nearly 100 per cent accuracy in speech transmission, at what would normally be an intolerably low signal-to-noise ratio.

Although much can be done to make speech communication possible, even in extremely intense noise (and only a few of the available methods have been mentioned), the discouraging feature of noise or unwanted sound is that noise at any audible level will tend to mask speech if the speech signal is not sufficiently intense. For this reason, it can be concluded that within broad limits the most deleterious effect of noise on man is its effect on speech communication.

Annoyance

A great deal of research and much discussion has been devoted to the question of how noise affects work performance—work not involving speech communication. Although noise can have an adverse influence on performance in certain rather special tasks requiring prolonged vigilance (Broadbent, 1951), most studies in this area have indicated that noise has little or no effect on efficiency (Kryter, 1950). Further, it appears that man can rather quickly adapt himself, provided the necessary speech communication requirements are fully met, to almost any noise environment. He may be paying for this adaptation by increasing his efforts, but so far this increased expenditure of energy, if any, has not been adequately measured.

Experiments and everyday observation clearly demonstrate that unexpected sounds or noise can be exceedingly obnoxious and cause violent startle responses. If such unexpected stimulation is excluded from consideration, is it possible to establish a criterion that will discriminate between an acceptable, "natural" auditory environment and a noise environment that is annoying? The idea that man has a "natural" auditory environment has a certain intuitive appeal as a basis for the construction of criteria for annoyance. Some of this appeal is lost, however, when one reflects on the following considerations. In the U. S. Navy, the maximum tolerable sound pressure level for berthing spaces is set in the lower octave band at 100 db above .0002 dyne/cm^2. Apparently, the ship's crew readily adapts to and sleeps in this amount of noise. On the other hand, it is a common experience that under some conditions the drip of a water faucet or the tick of a clock can be annoying.

It may be that noise is not always undesirable, and that there exist situations appropriate for the use of what might facetitiously be called "acoustic perfume." In most cases, this would be a low level noise with a broad frequenty spectrum.

594869 O - 61 - 13

Such a "perfume" might help induce sleep by making weak, intermittent sounds. It might also help radar scope watchers concentrate on their task, or even (by making conversation a little difficult) prevent secretaries from being distracted from their work.

The question of an annoyance threshold for noise was recently studied by Spieth (1956) at the Operational Applications Laboratory of the USAF Cambridge Research Center. Spieth asked military and civilian personnel to adjust the intensity level of various bands of noise until they were as loud as they could be without causing annoyance, assuming that the listener was to be immersed in the noise throughout his entire work day. The subjects were also asked questions about their present and previous work environments. On the basis of this information, and protocols taken after the tests, the subjects were divided into three groups. The group comprised those who were working or had worked in fairly intense noise, such as air crew members or industrial employees. These subjects were called "noise workers." Group 2 included those who had worked only in relatively quiet offices—called "quiet workers." The third group consisted of "quiet workers" who ignored the experimenter's instructions and did not imagine a definite work situation while making their observations.

The grand means of the judgments of the three groups, expressed in sound pressure levels relative to .0002 dyne/cm^2, were 88 db, 75 db, and 60 db, respectively. The differences among these means are statistically significant. It should be mentioned that the subjects were all less than 35 years of age and had no apparent hearing difficulties. A number of them were given audiometric tests. It seems unlikely, therefore, that the differences in annoyance thresholds for the "quiet" and "noise" workers were due to differences in hearing acuity.

Althouth there were wide individual differences among the subjects within each group, certain generalizations can be made. First, individuals who had been exposed in their work to aircraft or industrial noise had different thresholds of annoyance than persons who had not been so exposed. Second, the subjects who did not imagine a specific work situation—several reported that they attended to the noise as they would in a "psychophysical experiment"—wanted the noise at a somewhat lower level, in loudness terms, about 60 phons (Stevens, 1956). This threshold comes closest, perhaps, to representing annoyance judgments uncontaminated by other factors.

Why, it may be asked, do the "noise workers" tolerate 13 db more noise than the so-called "quiet workers"? Is it due to their having been "toughened" as the result of exposure to intense noise, or is it due, perhaps, to the fact that less speech communication is demanded in their jobs? Or, is it that they expect to use greater vocal effort, and are satisfied with a somewhat lower level of communication efficiency? Presumably, within broad boundaries, man's threshold of annoyance is strongly tempered, if not controlled, by the amount of auditory communication needed and the ease with which it can be carried out. However, some of the experimental data do suggest that in the absence of all needs to communicate or to perform a work task, people will indicate noise at a loudness level of about 60 phons as being at their threshold of annoyance.

On the basis of the kinds of experiments just described, and additional studies which cannot be discussed within the limitations of this report, criteria of acceptable noise levels for various kinds of military and industrial work spaces are being established. For example, it is proposed that a loudness level of 60 phons

be used as the criterion for acceptable noise, barely permitting reliable conversational and telephonic speech communication in executive offices. A criterion of no more than 55 phons would be desirable for exceptionally good conference-room and executive-office conditions, but certain types of office work can be accomplished without undue complaints in noise having loudness levels of the order of 77 phons. It has been found that the noise in many office buildings at military air fields equal or exceed this level.

References

Beranek, L. L. The design of speech communication systems. Proc. Inst. Radio Engrs., 1947, 35, 880-890.

Broadbent, D. E. The twenty dials and twenty lights test under noise conditions. M.R.C. Applied Psychology Research Unit Report No. 160, Unpublished Manuscript, Cambridge, England, 1951.

French, N. R., & Steinberg, J. C. Factors governing the intelligibility of speech sounds. J. acoust. Soc. Amer., 1947, 19, 90-119.

Kryter, K. D. The effects of noise on man. J. speech Dis., 1950, Monogr. Suppl. No. 1.

Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. J. exp. Psychol., 1951, 41, 329-335.

Stevems, S. S. Calculation of the loudness of complex noise. J. acoust. Soc. Amer., 1956, 28, 807-832.

Spieth, W. Annoyance threshold judgments of bands of noise. J. acoust. Soc. Amer., 1956, 28, 872-877.

Summary

Developing communication systems which will perform satisfactorily under conditions of moderate to intense noise is an important military problem. The present paper describes recent research on speech communication under noise and points out how the results can be applied to the solution of practical military problems. Research on the effect of noise on human work efficiency is also described.

A measure of the intelligibility of speech for any given communication system can be obtained by computing an "Articulation Index" designed to reflect the

overall signal-to-noise ratio for the frequency spectrum of speech sounds. The Index is computed by dividing the spectrum into 20 frequency bands of equal importance to speech intelligibility, and averaging the individual signal-to-noise ratios computed in terms of decibels for each band.

This method of calculation assumes that the masking of a portion of the speech spectrum by noise has the same effect on intelligibility as the removal of that portion by frequency distortion or filtering. Although recent studies have shown that special masking effects occur when narrow-band noise is present, the method has been shown to yield results sufficiently accurate for practical design purposes under the noise and frequency-distortion conditions usually present in military communication systems.

The Articulation Index best predicts understandability for connected discourse. Figure 1 shows the percentage of different types of messages (syllables, words, and sentences) found under laboratory testing conditions to be correctly understood at various signal-to-noise ratios. Figure 2 shows a work sheet used in applying the Articulation Index method to an aircraft interphone system.

It would appear that, to improve speech intelligibility for communication systems having unacceptable Articulation Indices, all that is necessary is to specify what alterations in the noise or speech spectrum are required to raise understandability to a satisfactory level. However, reducing the noise or raising the speech level may prove impractical or too costly, and increases of speech-signal intensity must stop short of the point at which ear damage results.

An alternative approach is to design messages with human perceptual capabilities in mind. One method, based on the finding that the understandability of a message rises with its redundancy, proceeds through increasing message length. Under certain circumstances, however, this approach may prove too inefficient in terms of time. A second method involves minimizing the size of the "vocabulary" from which the messages are constructed. It has been shown that message reception at a given signal-to-noise varies inversely with vocabulary size when the listener has been familiarized with the vocabulary being used (see Fig. 3). The U. S. Air Force has recently applied this inexpensive method to improvement of its Air Defense communication network with salutary results.

Turning to a consideration of the effects of environmental noise on work efficiency, it is noted that noise can have an adverse influence on performance in tasks requiring prolonged vigilance, but that little or no measurable effects on efficiency have been found for work of other types.

A study is described in which the investigators sought a means of differentiating "acceptable" noise environments from those which are annoying. This problem is complicated by the fact that human subjects can work, and even sleep, in very intense noise, but may, at other times, be disturbed by extremely weak sounds, such as the dripping of a faucet.

The research in question involved asking military and civilian personnel to adjust the intensity of various bands of noise until they reached the maximum level which could be borne without annoyance throughout an entire work day. The subjects were then divided into three subgroups, and comparisons were made of the average annoyance thresholds of those subjects who had: (a) had experience working in very noisy environments, (b) worked only in relatively quiet offices, and (c)

worked in quiet conditions, but ignored the experimenter's instructions to make their threshold estimations while imaging a definite work situation. All of the subjects were under 35 years of age and had no apparent hearing difficulties.

Analysis of the results showed that although individual differences were marked, those subjects who had previously been exposed to noisy work environments tended on the average to have higher annoyance thresholds than subjects previously employed in relatively quiet environments. It was also found that as a group, the subjects who had not imagined a definite work situation when making their observations, had the lowest expressed tolerance for noise. The estimates of this group probably best represent "pure" annoyance judgments uncontaminated by other factors.

Possible explanations of the findings for the first two groups are: (a) that experience with noisy work environments results in auditory adaptation, (b) that less speech communication generally is required on noisy jobs, and (c) that levels of acceptance regarding the amount of vocal effort required for speech communication under noise are raised by previous experience. It is hypothesized that annoyance thresholds are tempered by the amount of auditory communication required in various noise environments, and by the ease with which such communication can be carried out.

Studies of this kind have resulted in the establishment of acceptable noise levels for various kinds of military and industrial work. It has been found that the noise in office buildings at U. S. military air fields often equals or exceeds acceptable levels.

Résumé

Un problème militaire important est le développement de systèmes de communication donnant satisfaction dans des conditions de bruit modéré à intense. Le présent rapport expose des études récentes portant sur la communication verbale dans des conditions de bruit, et montre comment ces résultats peuvent être appliqués à la solution de problèmes militaires pratiques. Une étude de l'effet du bruit sur le rendement professionnel de l'homme est également exposée.

On peut mesurer l'intelligibilité de la parole pour tout système de communication donné en calculant un "Indice d'Articulation" destiné à refléter le rapport total signal:bruit pour le spectre de fréquences des sons produits par la parole. On calcule l'Indice en divisant le spectre en 20 bandes de fréquence d'importance égale quant à l'intelligibilité de la parole, et en prenant la moyenne entre rapports individuels signal:bruit calculés en décibels pour chaque bande.

Cette méthode de calcul suppose que le fait de masquer une partie du spectre de la parole par du bruit a le même effet sur l'intelligibilité que l'enlèvement de cette partie par distorsion ou filtrage des fréquences. Bien que des études récentes aient démontré qu'il se produit des effets masquants spéciaux quand on se trouve en présence de bruit à bande étroite, la méthode a donné des résultats suf-

fisamment précis pour être appliquée à la pratique dans les conditions de bruit et de distorsion de fréquences que l'on rencontre habituellement dans des systèmes de communication militaires.

L'Indice d'Articulation convient le mieux pour prédire la compréhensibilité de discours suivi. La Figure 1 montre le pourcentage de différentes espèces de messages (syllabes, mots, et phrases) que l'on a constaté être correctement compris, dans les conditions d'essai de laboratoire, pour différents rapports signal: bruit. La Figure 2 représente une feuille de travail utilisée en calculant l'Indice d''Articulation pour un système téléphonique intérieur dans un avion.

Il semblerait que, pour améliorer l'intelligibilité de la parole pour des systèmes de communication avec Indices d'Articulation inacceptables, il suffit de spécifier quels changements sont nécessaires dans le spectre de bruit ou de parole pour porter la compréhensibilité à un niveau satisfaisant. Cependant, la réduction du niveau du bruit ou l'augmentation de celui de la parole peut se révéler irréalisable ou trop coûteuse, et des augmentations de l'intensité de la parole doivent s'arrêter avant de causer des dommages à l'oreille.

On peut également y arriver en composant des messages tenant compte des capacités de perception humaines. Une méthode, basée sur la constatation que la compréhensibilité d'un message augmente avec sa redondance, consiste à augmenter la longueur du message. Dans certains cas, cependant, cette méthode peut être trop inefficace en fonction du temps. Une seconde méthode consiste à diminuer l'étendue du "vocabulaire" dont les messages sont composés. On a constaté que la réception de messages à un rapport signal:bruit donné varie en proportion inverse de l'étendue du vocabulaire, quand l'auditeur s'est familiarisé avec le vocabulaire utilisé (voir Fig. 3). L'Armée de l'Air américaine a récemment appliqué cette méthode peu coûteuse pour améliorer son réseau de communication pour la défense anti-aérienne, et en a obtenu de très bons résultats.

Examinant les effets des bruits ambiants sur le rendement du travail, il fait remarquer que les bruits peuvent avoir une mauvaise influence sur l'efficacité pour des tâches exigeant une attention prolongée, mais qu'on a constaté peu ou pas d'effets mesurables sur des travaux d'autres genres.

Il décrit une étude dans laquelle ses collègues ont cherché un moyen de distinguer des bruits "acceptables" des bruits gênants. Ce problème se complique du fait que des sujets humains peuvent travailler et même dormir au milieu de bruits très intenses, mais peuvent, à d'autres moments, être gênés par des sons extrêmement faibles tels qu'un robinet qui coule goutte à goutte.

Dans la recherche en question, on a demandé à du personnel militaire et civil de régler l'intensité de diverses bandes de bruits jusqu'à ce qu'ils atteignent le niveau maximum qui peut être supporté sans gêne pendant toute une journée de travail. Les sujets ont alors été divisés en trois groupes et on a comparé les seuils moyens de gêne des sujets qui (a) avaient l'habitude de travailler dans des milieux bruyants, (b) n'avaient travaillé que dans des bureaux relativement calmes, et (c) avaient traveillé dans des conditions calmes mais n'ont pas tenu compte des instructions de l'expérimentateur d'estimer les seuils en imaginant qu'ils se trouvaient dans une certaine situation de travail. Tous les sujets étaient âgés de moins de 35 and et n'éprouvaient aucune difficulté apparente de l'ouïe.

L'analyse des résultats a montré que, bien que les différences entre individus fussent nettes, les sujets qui avaient été exposés auparavant à des milieux de travail bruyants présentaient généralement des seuils de gêne plus élevés que les sujets travaillant antérieurement dans des milieux relativement calmes. On a constaté aussi que, pris dans leur ensemble, les sujets qui n'avaient pas imaginé une situation de travail déterminée en faisant leurs observations exprimaient la tolérance au bruit la plus basse. Les estimations de ce groupe représentent probablement le mieux des jugements de gêne "purs" non influencés par d'autres facteurs.

Les constatations pour les deux premiers groupes peuvent s'expliquer comme suit: (a) il résulte de l'expérience de milieux de travail bruyants une adaptation auditive, (b) le besoin de communication orale est généralement moindre dans des travaux bruyants, et (c) les niveaux d'acceptation en ce qui concerne l'intensité d'effort vocal nécessaire pour la communication orale dans le bruit sont augmentés par l'expérience préalable. On suppose que les seuils de gêne sont modérés par la quantité de communication auditive nécessaire dans divers milieux de bruit et par la facilité avec laquelle une telle communication peut être effectuée.

Des études de cette espèce ont mené à l'établissement de niveaux de bruit acceptables pour divers types d'activités militaires et industrielles. On a constaté que le bruit dans les immeubles de bureaux des champs d'aviation militaires aux Etats-Unis est souvent égal à ou dépasse les niveaux acceptables.

Commentary by Discussant

Dr. R. J. H. Kruisinga (The Netherlands):

In commenting on Dr. Kryter's excellent paper, I should like to give primary attention to the section dealing with speech communication under noise. The author has mentioned a number of solutions to the problem of how to alleviate the effects of noise on message intelligibility. In addition to these methods, attention can be drawn to the possibilities offered by: (a) alteration of the phonetic composition of the message, (b) the use of selective, non-linear amplification, (c) use of the whispered voice, and (d) special training for communications operators.

First, let us consider the results of research on phoneme analysis. As you are no doubt aware, pronunciation of various vowels and consonants produces distinctive sound patterns of characteristic frequencies and intensities. The understandatility of these patterns at different signal-to-noise ratios has been found, not unexpectedly, to vary. For example, at a signal-to-noise ratio of five decibels the consonant "t" is correctly recognized only 40 per cent of the time, but the consonant "r" is understood about 90 per cent, and the vower "a" is understood nearly 100 per cent of the time.

Given data of this kind, it is possible to calculate a "phoneme articulation index" for each individual vowel and consonant at various signal-to-noise ratios. By further determining which phonemes tend to be most often confused with each other at given ratios, one can formulate an objective "phoneme confusion schema"

which can be used as a guide in the composition of messages. The implications of these principles for military code construction are clear: in drawing up a message set, attempts should be made to choose code words of optimal phonetic composition for the signal-to-noise-ratio conditions in which the code will be used.

Message intelligibility under noise can also be improved through the use of whispered speech. Since the sound patterns of whispered and spoken phonemes are different, reception under certain conditions is better for whispered than for spoken messages. Moreover, this advantage can be increased through the use of special, non-linear amplification methods—a procedure which simultaneously permits the filtering out of unwanted background noise.

Interestingly enough, it has also been found that applying selective amplification through two different channels to the right and left ears produces further increases in understandability. Why this should be so is still a point for research, and the theoretical consequences may prove very interesting.

Finally, we should not overlook the gains to be derived from special training of communications operators. By training operators on certain frequency bands, and drawing their attention to the specific difficulties and phoneme confusion characteristics of those bands, better Articulation Indices can be achieved. This approach involves the same principles as those currently used in working with hard-of-hearing patients.

In conclusion, I should like to stress the need for additional research. We must try to discover which elements in spoken and whispered speech are important for its recognition as such, and which elements are merely noise and exert a masking effect. Given these facts, we should be in a much better position to maximize the amount of information transmitted at high levels of noise.

General Discussion

Mr. Donald E. Broadbent (U. K.):

Dr. Kryter's paper has suggested three points which I should like to submit for discussion. The first concerns the effects produced on communication by exposure to noise which is not present when the communication is made. I am referring to the kind of deafness which sometimes produces not only a rise in threshold, but also an abnormal loudness-intensity function. On theoretical grounds, one might question whether the Bell Telephone calculation system would yield trustworthy results for individuals with abnormal loudness curves. If anyone has attempted to utilize the system with subjects deaf from exposure to noise, I should be greatly interested to hear of the results.

The second point concerns the possibility of assessing communication channels by measures other than intelligibility. It has been shown that removal of the information in a portion of the spectrum by filtering produces the same effects on understandability as does the removal of that information by distortion, i.e., by

turning the signals into noise. However, we have found that the two communication channels which result are not equally easy to listen to: one requires more concentration and produces more serious effects on the performance of a simultaneous visual task. This suggests the existence of some additional factors in communication to which the intelligibility index is not sensitive.

My final point relates to the problem of the effects of noise on efficiency. It is undoubtedly true that performance on many tasks is uninfluenced by noise. But, there is an increasing number of experiments which do show effects—curious ones which are difficult to understand. For example, in one recent study, a group of subjects was required to carry out a task in intermittent noise. As in many other experiments, it was found that the average level of performance was the same under these conditions as in a noise-free environment. However, after each burst of noise, a brief period (about 30 seconds) of reduced efficiency occurred. Another illustration is a recently completed experiment in which it was shown that working in noise on one day produced an effect on work carried out in quiet the next day. I have no explanation of these findings and would appreciate hearing some comments on them.

Reply

Dr. Karl Kryter:

I was very much interested in Dr. Kruisinga's remarks about whispered speech. This appears to be an important development which should be carefully and thoroughly studied.

As regards phoneme analysis, I am in hearty agreement that more can often be done to combat noise through language engineering and special training than through increasing the power level of amplifiers or the frequency response of earphones. Studies of human perception form a vital part of systems research, since they often suggest ways of obtaining satisfactory results without making changes in our equipment.

I should answer Mr. Broadbent's question concerning whether the Articulation Index will work with deaf subjects by saying that it depends on the type of deafness. For persons with otosclerosis, or conductive-type deafness, audiograms plotted in the same way as the noise spectrum yield results very similar, I believe, to those obtained by the AI method. For individuals with nerve-type deafness due to noise, however, it is difficult to say. Work done on the limits of intelligibility for exposure-deaf subjects has revealed marked individual differences, and one would expect that different types of deficits would affect AI results in different ways.

As regards the effects of noise on work efficiency, it is true that these are sometimes curious. They are not, however, always or necessarily deleterious. Apparently, some types of noise tend to isolate the individual from his environment by masking distracting auditory stimuli, and, therefore, help to increase efficiency.

MILITARY SUCCESS OF MENTALLY SUBNORMAL SOLDIERS

Eljas Siro
Institute of Psychology, University of Helsinki
Finland

Since 1901, special teaching has been provided in Finland for children who are unable to follow ordinary class work because of mental subnormality. At present, this special training is given to all children with IQ's between about 60 and 80. During the early stages of the program the transfer to a special class was made at the suggestion of the child's teacher, but soon after 1930 the use of intelligence tests was adopted. This method of extimating the school aptitude of mentally subnormal children did not become fully established until 1946, however, when standardization of the Terman-Merrill test was completed in Finland.

In the armed forces mental subnormality is considered to reduce or handicap fitness for service. At present, mildly subnormal men (IQ-60 to 80) are assigned to manual labor, while those with IQ's of less than 60 are freed from the service entirely. All men with IQ's greater than 80 are accepted for active service. When carrying out fitness classifications the Army authorities do not have at their disposal any information as to whether the conscripts have attended special school classes.

Study 1

During World War II it was found in many countries that the lower limit of intelligence assumed to be necessary for front line service had been set too low, and the cut-off level for assignment to such duty had to be raised accordingly. The purpose of the study was to determine how the military performance of mentally subnormal men (defined as those who had attended special school classes) compared with that of normally intelligent men of the same age during the wartime periods of 1939-40 and 1941-45. The research was begun in 1949. Since, at that time, it was no longer possible to get in touch with the men in question, the analysis was based on available documents and records.

The study was limited to men who had attended school in Helsinki. The experimental group (subnormal subjects) consisted of men born during the period 1910-1919 who, because of their poor school achievement, had been transferred to special classes. Only two of the subjects in this group had later been returned to regular school. It should be noted that since transfers to special classes were not based on intelligence tests at the time these men attended school, it is possible that some of the experimental subjects may have been intellectually normal (or subnormal) psychopaths.

The control group consisted of men who had attended retular school classes. The sample was composed by randomly selecting from the available enrollment lists a matching group of the same age and size for each age group of the experimental sample.

The analysis of the data consisted of comparing the experimental and control groups with respect to the following variables:

1. Classification as to fitness for service made by the military authorities: (a) at the time of enrollment, and (b) at the end of the war;

2. Type of wartime duties performed;

3. Convictions for criminal offenses.

Military fitness classifications are always carried out by physicians. Thus, with the exception of cases referred at military hospitals during the war, none of the subjects received detailed psychiatric examinations.

A basic change in the regulations governing fitness classifications was made in 1935. Prior to that time it was customary to reject all men considered by the physician as too mentally defective to be able to profit from training, or to carry out military duties. In 1935, a more explicit grading and assignment procedure was adopted. The new system specified that only "idiots" and "imbeciles" (IQ below 55/60) were to be considered unfit for any kind of service, and that "debiles" (IQ between 55/60 and 75/80) were to be accepted into the service and assigned to labor duty. Half the men in the experimental group of the present study were classified according to the pre-1935 regulations (established in 1923), and the other half were classified according to the system adopted in 1935. The estimations of mental ability were not made on the basis of intelligence test results either before or after 1935, and, as previously mentioned, information on special-class attendance was not available to the physicians who carried out the evaluations.

Results

1. Fitness classifications at time of enrollment. Table 1 shows the number of experimental and control subjects who, at the time of enrollment, were: (a) classified as fit for front line duty, (b) classified as fit for labor duty, and (c) rejected from service. Separate data are given for subjects evaluated in accordance with the regulations in effect before and after 1935 (men born during the periods 1910-1914 and 1915-1919, respectively). It will be seen that the classification of the experimental group differs significantly from that of the control group for those subjects called up prior to 1935, whereas for the post-1935 conscripts the difference between the classification distributions of the two groups is not statistically significant.

2. Fitness classifications at war's end. Data showing the classifications carried out by physicians at the end of the war are presented in Table 2. These evaluations probably constitute better indicators of actual fitness for service than those made at the time of enrollment. Serious attempts were made to keep information on the troops' physical and mental condition up to date, and it may be assumed that repeated checks of the classifications were made during the wartime service period.

Comparison of the prewar and postwar classification figures in Tables 1 and 2 shows that the enrollment classifications made according to the 1923 regulations remained virtually unchanged at the end of the war. The corresponding differences between the experimental and control group distributions are highly significant in both tables. The postwar classifications of the experimental and control subjects evaluated in terms of the 1935 system, however, are considerably different from

TABLE 1
FITNESS CLASSIFICATION OF EXPERIMENTAL AND CONTROL GROUPS AT TIME OF ENROLLMENT FOR WARTIME SERVICE

	Regulations of 1923			Regulations of 1935		
	Experimental group	Control group	Total	Experimental group	Control group	Total
Front line service	43	69	112	61	78	139
Labor service	16	16	32	2	5	7
Rejected	21	3	24	20	14	34
Total	80	88	168	83	97	180
	$x^2 = 19.2$	df = 2		$x^2 = 2.4$	df = 2	
	p = <.001			p = <.30		

TABLE 2
FITNESS CLASSIFICATION OF EXPERIMENTAL AND CONTROL GROUPS AT END OF WAR

	Regulations of 1923			Regulations of 1935		
	Experimental group	Control group	Total	Experimental group	Control group	Total
Front line service	37	60	97	33	58	91
Labor service	19	24	43	21	31	52
Rejected	24	4	28	29	8	37
Total	80	88	168	83	97	180
	$x^2 = 20.7$	df = 2		$x^2 = 19.9$	df = 2	
	p = <.001			p = <.001		

those made at enrollment. At the time of entry into service approximately equal proportions of the two post-1935 groups had been considered capable of front line duty, but by the end of the war a much greater reduction had occurred in the number of subnormal than in the number of normal subjects considered fit for such service. Thus, the difference between the experimental and control classifications, negligible at the time of induction, is highly significant at war's end. (It should be noted parenthetically that the proportion of cases reclassified because of wounds or other casualties was approximately equal in the two groups.)

It would appear, therefore, that of the two sets of classification regulations the one in effect before 1935 possessed greater value as a method of estimating the military capability of men who had attended special classes at school. This may have been due to the essentially "common sense" nature of the earlier procedure. Although the regulations established in 1935 provided a more specific classification system, no provisions were made for securing expert advice in connection with their application.

3. Type of wartime duties performed. When the wartime service records of the subjects were examined, it was found that there were only three non-commissioned officers (two sergeants and one sergeant-major) in the experimental group, and at least one of these had subsequently been demoted to the rank of common soldier near the end of the war. In contrast, the control group included 29 non-commissioned officers, 24 commissioned officers, and 6 military officials.

Tabulation of the duties performed by the troops showed that 53 per cent of the experimental and 9 per cent of the control subjects had been transferred to labor service. Even if only those control subjects who did not proceed beyond elementary school are considered, the difference in favor of the control group remains statistically significant.

4. Criminal offenses. Criminality is an important indicator of lack of adjustment. When all convictions for criminal offenses committed prior to, during, and after the period of military service are considered, it is found that 68 of the experimental and 29 of the control subjects had at some time been sentenced to probational or non-probational imprisonment prior to 1949. As is apparent from Table 3, the difference between the groups is highly significant.

Analysis of the nature of the offenses committed during the period of service showed that those of the experimental subjects consisted for the most part of running away, desertion, and exceeding leave of absence. In the control group, on the other hand, only 50 per cent of the offenses were of this type. The other 50 per cent involved theft, intoxication, personal assault, etc. The difference between the two groups with respect to the type of crime committed is significant beyond the .001 level of confidence ($x^2 = 12.49$).

It is known that a considerable portion of the criminal population falls in the IQ range 70 to 90. The fact that the experimental and control groups differed significantly with respect to type of crime committed lends support to the hypothesis that the experimental subjects had been transferred to special classes primarily for reasons of intellectual subnormality, rather than for reasons of psychopathy.

Study 2

In order to compare the above results with those obtained at a later date under peacetime conditions, an additional study was carried out utilizing subjects who had been born in 1932 and called up for military service in 1951. The experimental

TABLE 3
COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS WITH RESPECT TO CRIMINAL MILITARY AND CIVILIAN OFFENSES COMMITTED PRIOR TO 1949

	Experimental group	Control group	Total
Sentenced	68	29	97
Not sentenced	107	167	274
Total	175	196	371
	$x^2 = 27.1$	df = 1	p = <.001

group of this study consisted of all men entered on the 1951 enrollment list of the military district of Helsinki who had attended special school classes. The control group, consisting, as before, of men who had attended regular school classes, was obtained via random selection from the same enrollment list. As in the previous study, information on special class attendance was not available to those charged with carrying out military fitness classifications.

When the experimental and control groups were compared with respect to fitness classification (a) at the time of enrollment, and (b) after eight to eleven months of service, no significant differences were found in either case (see Table 4). However, analysis of the type of duties carried out revealed a marked difference between the groups. In actuality, 41 per cent of the experimental as opposed to 2 per cent of the control subjects were employed on labor assignments during the training period. These percentages do not markedly differ from those for the wartime groups.

A final comparison was based on the ratings made by training officers of the subjects' success in training. Table 5 shows that in terms of this criterion also, the group of experimental subjects proved to be significantly inferior to the group of normal controls.

TABLE 4
FITNESS CLASSIFICATION OF CONSCRIPT GROUPS CALLED UP IN 1951

	At Enrollment			After 8-11 Months of Service		
	Experimental group	Control group	Total	Experimental group	Control group	Total
Front line service	37	38	75	33	30	63
Labor service	1	1	2	1	2	3
Rejected	0	2	2	2	2	4
Total	38	41	79	36	34	70

TABLE 5
RATED SUCCESS OF EXPERIMENTAL AND CONTROL SUBJECTS DURING PEACETIME MILITARY TRAINING

	Experimental group	Control group	Total
Good	19	27	46
Moderate or poor	15	6	21
Total	34	33	67

$x^2 = 5.9$ df = 1 $p = <.02$

Summary and Conclusions

The purpose of the research was to compare the military success of intellectually normal and subnormal Finnish conscripts under wartime and peacetime conditions. To this end, groups of men who had attended special classes for the mentally subnormal were compared with groups of the same age who had attended regular school classes. Comparisons were made with respect to: (a) fitness classifications made at the time of enrollment and at the end of specified periods of service, (b) type of military duties performed, (c) convictions for criminal offenses, and (d) ratings of success in (peacetime) training.

The results showed that, in general, the subnormal subjects were markedly inferior to the normal conscripts on the various performance criteria studied. It was also found that the classification regulations introduced in 1935 were less efficient than those established in 1923 for predicting the military capability of special-class subjects. A considerable readjustment of the enrollment classifications of the subnormal conscripts had to be made in the light of their wartime performance. Further evidence for the inadequacy of the present classification system derives from the finding that a significantly larger number of subnormal than of normal subjects had actually been employed on labor assignments during peacetime training, despite the fact that the official fitness classifications of the two groups did not significantly differ, either at the beginning or at the end of the training period. It can be concluded that a reduction of classification errors could be achieved by assigning all conscripts who have attended special school classes directly to labor service at the time of induction.

Summary

The purposes of the research were: (a) to compare the military success of intellectually normal and subnormal Finnish conscripts under wartime and peacetime conditions, and (b) to determine which of two military fitness classification systems yields better predictions of performance for men of subnormal intelligence.

Study 1 was based on records of performance in service during World War II. The investigation was limited to men who had attended school in Helsinki. The experimental sample (subnormal subjects) consisted of men born during the period 1910-1919 who, because of their poor school achievement, had been transferred to special classes for the mentally defective. (Special teaching is provided in Finland for all children having tested IQ's between 60 and 80.) The sample of normal (control) subjects, matched for age with the experimental sample, was composed by random selection from the available enrollment lists.

The analysis of the data consisted of comparing the two groups of subjects with respect to: (a) fitness-for-service classifications at the time of enrollment and at the end of the war, (b) type of wartime duties performed, and (c) confictions for criminal offenses.

Finnish military fitness classifications are carried out by regular Army physicians. The evaluations are not based on the results of intelligence tests, and the physicians are not provided with information on whether the conscripts have attended special classes at school.

Prior to 1935, it was customary to reject all men considered by the examining physician as too mentally defective to perform successfully in the service. In 1935 more precise classification regulations were adopted. These specified that "idiots" and "imbeciles" (IQ below 60) were to be rejected from the service, "debiles" (IQ 60 to 80) were to be assigned to military labor duty, and men with IQ's above ip were to be classified as fit for active service.

Half of the experimental and control subjects of the study were classified according to the pre-1953 regulations, and half were classified according to the post-1935 procedure. Experimental-control comparisons with respect to prewar and postwar fitness classifications were made separately for the two groups.

The results of the analysis were as follows:

1. The classification of the experimental group at the time of enrollment differed significantly from that of the control group for the pre-1935 subjects, but the intergroup difference was not statistically significant for subjects evaluated according to the regulations of 1935 (Table 1).

2. The original classification distributions for the experimental and control subjects evaluated prior to 1935 remained virtually unchanged at the end of the war. For subjects evaluated in terms of the 1935 system, however, it was found that a large number of the subnormal conscripts, classified at enrollment as fit for front line duty, had to be transferred to labor assignments during the course of the war. The difference between the experimental and control distributions, negligible at the time of induction, was thus highly significant at war's end (compare Tables 1 and 2).

3. Tabulation of wartime duties performed showed that a much larger percentage of experimental than of control subjects had been employed on labor assignments. The difference between the groups remained statistically significant, even when only those control subjects who had not progressed beyond elementary school were used as the basis of comparison.

4. A significantly larger number of the subnormal than of the normal subjects had been sentenced to prison for military and/or civilian criminal offenses prior to the time of the study (Table 3). Military crimes committed by the subnormal subjects consisted primarily of desertion or exceeding leave of absence. The control subjects committed a significantly larger number of offenses involving theft, drunkenness, or violence.

Study 2 compared the success of normal and subnormal conscripts in peacetime military training. The experimental and control groups consisted, respectively, of men entered on the 1951 military enrollment list for the district of Helsinki who: (a) had, and (b) had not attended special school classes for the mentally defective.

The differences between the two groups with respect to military fitness classification made: (a) at the time of enrollment, and (b) after eight to eleven months

of service were not statistically significant (Table 4). However, analysis of the duties actually performed during the training period showed that 41 per cent of the experimental, as compared with 2 per cent of the control subjects, had, in fact, been employed on labor assignments. Comparison of the groups with respect to ratings of success in training showed that the subnormal subjects were also significantly inferior on this performance criterion (Table 5).

In discussing the results, the author points out that the earlier, less explicit classification regulations appear to be more efficient than those adopted in 1935 as a means of estimating the military capability of intellectually subnormal conscripts. A large number of the subnormal subjects evaluated in terms of the current regulations had to be reclassified or reassigned in the light of their in-service performance. It is suggested that a reduction of classification errors could be achieved by assigning all conscripts who have attended special school classes directly to labor service at the time of induction.

Résumé

L'étude avait pour buts: (a) de comparer le succès militaire de recrues finnoises d'un niveau intellectuel normal et inferieur à la normale dans les conditions de guerre et de paix, et (b) de déterminer lequel des deux systèmes de classification d'aptitude militaire permet le mieux de prévoir le rendement d'hommes ayant une intelligence inférieure à la normale.

L'étude n° 1 fut basée sur des documents indiquant le rendement en service pendant la IIè Guerre Mondiale. La recherche s'est limitée à des hommes qui avaient fait leurs études à Helsinki. Le groupe expérimental (sujets au-dessous de la normale) se composait d'hommes nés pendant la période 1910-1919 qui, vu leurs résultats médiocres a l'école, avaient été transférés dans des classes spéciales pour déficients mentaus. (Un enseignement spécial est fourni en Finlande pour tous les enfants ayant un quotient d'intelligence compris entre 60 et 80.) Le groupe de sujets normaux (de contrôle) fut composé en choisissant au haxard, sur les listes de recrutement, un nombre égal d'hommes du même âge.

L'analyse des données consistait à comparer les deux groupes de sujets en ce qui concerne: (a) les classifications d'aptitude au service au moment du recrutement et à la fin de la guerre, (b) le genre de fonctions effectuées pendant la guerre, et (c) les condamnations pour délits.

Les classifications d'aptitude militaire sont établies en Finlande par les médecins militaires. Les évaluations ne sont pas basées sur les résultats de tests d'intelligence, et les médecins ne savent pas si les recrues ont suiti des cours spéciaux a l'école.

Avant 1935, il était d'usage de refuser tous les hommes considérés par le médecin qui les examinait comme trop déficients au point de vue mental pour s'acquitter convenablement de leurs tâches. En 1935, des règlements de classification plus précis ont été adoptés. Ils spécifaient que les "idiots" et les "imbéciles"

(intelligence inférieure à 60) doivent être considéres comme inaptes au service, les "décibiles" (intelligence de 60 à 80) doivent être affectés au service militaire du travail, et les hommes d'une intelligence supérieure à 80 doivent être considérés comme aptes au service actif.

La moitié des sujets d'expérience et de contrôle de l'étude fut classée suivant les règlements d'avant 1935, et la moitié fut classée suivant la méthode d'après 1935. Les comparaisons entre les sujets d'expérience et de contrôle concernant les classifications d'avant-querre et d'après-guerre ont été faites séparément pour les deux groupes.

L'analyse a donné les résultats suivants:

1. La classification du groupe expérimental au moment du recrutement a différé d'une manière significative de celle du groupe de contrôle pour les sujets d'avant 1935; mais la différence entre les groupes n'a pas été statistiquement significative pour les sujets évalués suivant les règlements de 1935 (Table 1).

2. Les répartitions originales de classification pour les sujets d'expérience et de contrôle évalués avant 1935 sont restées virtuellement inchangées à la fin de la guerre. Pour les sujets évalués selon le système de 1935, cependant, on a constaté qu'un grand nombre des recrues au-dessous de la normale classées au moment du recrutement comme aptes à aller en première ligne, ont dû, pendant la guerre, être versées au service du travail. La différence entre les répartitions d'expérience et de contrôle, negligeable au moment du recrutement, a donc été hautement significative à la fin de la guerre (comparer les Tables 1 et 2).

3. L'analyse des fonctions effectuées pendant la guerre a montré qu'un pourcentage beaucoup plus grand de sujets d'expérience que de sujets de contrôle avait été employé au service du travail. La différence entre les groupes est restée statistiquement significative même quand on ne prenait comme base de comparaison que les sujets de contrôle qui n'avaient pas dépassé l'enseignement primaire.

4. Un nombre beaucoup plus grand de sujets au-dessous de la normale que de sujets normaux avaient été condamnés à des peines de prison pour des délits militaires et/ou civils avant l'époque de l'étude (Table 3). Les délits militaires commis par les sujets au-dessous de la normale consistaient surtout en cas de désertion et de prolongation de permissions. Les sujets de contrôle avaient commis un nombre plus grand de délits tels que vol, ivrognerie ou violence.

L'étude n° 2 compara le succès de recrues normales et au-dessous de la normale pendant le service militaire en temps de paix. Les groupes d'expérience et de contrôle se composaient, respectivement, d'hommes inscrits sur la liste de recrutement militaire de 1951 pour le district d'Helsinki qui: (a) avaient, et (b) n'avaient pas suivi de cours spéciaux pour déficients mentaux.

Les différences entre les deux groupes quant aux classifications d'aptitude militaire faites: (a) au moment du recrutement, et (b) après huit à onze mois de service, n'étaient pas significatives au point de vue statistique (Table 4). Cependant, l'analyse des fonctions réellement effectuées pendant la période de service a montré que 41 pour cent des sujets d'expérience, contre 2 pour cent des sujets de contrôle, avaient en fait été employés comme hommes de peine. La comparaison des groupes quant aux cotes de succès faites par les autorités militaires a montré que les sujets au-dessous de la normale étaient très inférieurs aussi d'après de critère de rendement (Table 5).

Au cours de sa discussion des résultats, on signale que les règlements de classification précédents, bien que moins explicites, semblent être plus efficaces que ceux adoptés en 1935 comme moyen pour évaluer les aptitudes militaires de recrues intellectuellement au-dessous de la normale. Un grand nombre des sujets inférieurs à la normale, évalué en fonction des règlements courants, a dû être reclassé ou réaffecté selon leur rendement en service. On pense que les erreurs de classification pourraient être réduites en affectant toutes les recrues qui ont suivi des cours speciaux directement au service du travail au moment de l'admission.

Commentary by Discussant

Dr. H.-K. Knoepfel (Switzerland):

Dr. Siro has presented us with some valuable information on the military capabilities of mental defectives. This problem is a very important one in the Armed Forces. Asice from the fact that they are difficult to train and often cause social and disciplinary problems, mentally deficient soldiers can constitute a real danger to their units in combat.

In dealing with individual cases, however, we should exercise caution when making diagnostic decisions. We know from clinical experience that there exist conditions of neurotic inhibition and other affective disorders which may interfere with intellectual functioning. Such cases, if properly diagnosed, might be saved for military service by appropriate psychotherapy. I have personally seen many cases of low or borderline intelligence who might have been capable of active duty, had their performance not been further impaired by neurotic difficulties which rendered them as little useful to their units as if they had been true mental defectives. Since individuals of this type may be amenable to treatment, it is important to differentiate as clearly as possible between mental defect and poor intellectual performance due to affective disorders.

The military psychologist and psychiatrist should, therefore, not stop at mere selection. He should also give attention to therapeutic possibilities.

General Discussion

Dr. Joseph Zubin (U. S. A.):

I am wondering whether we haven't tended to over-rate the role played by measured intelligence in military effectiveness. It should not be forgotten that mental deficiency is basically a socially defined condition, and that many people with measured I.Q.'s of 50, and even below, have been found capable of getting along under limited conditions in the community. This and other evidence suggest that the measuring instruments commonly used to separate the mentally "normal" from the mentally "defective" may not be as valid as is often assumed. It is well known, for example, that individuals who are educationally deprived may show up as mentally deficient on verbal intelligence tests, but obtain scores well within the normal range on such tests as the Porteus Maze. Today, military manpower needs are such that we should exercise extreme caution with respect to the measurement criteria in terms of which mental deficiency is defined. Current selection procedures may not be optimal, and it is conceivable that research on the screening potentialities of perceptual, psychomotor, and psychophysical tests, for example, would show that many so-called mental defectives are capable of performing satisfactorily in certain military occupations.

At this time I should like to point to the research possibilities inherent in the great masses of test data collected each year in all countries having military drafts. The provisions for testing which already exist in these countries appear to provide an excellent opportunity for collecting comparative data on the international incidence of various types of mental defect and mental illness. Given a standard test battery and a central research committee, information could be obtained on many clinical questions of great practical and theoretical importance.

Reply

Mr. Eljas Siro:

I should like to thank the discussants for their most interesting contributions to the problem treated in my paper. Although the handling of mental defectives may vary in different countries according to local conditions, the problem itself is of equal importance everywhere.

THE DISCRIMINABILITY OF SIMULATED VISUAL SIGNALS

J. Alfred Leonard
Applied Psychology Research Unit, Medical Research Council
Cambridge, England

A recurrent problem in experimental psychology is that of specifying the stimulus material and varying it along meaningful dimensions. Recently, a novel approach has been adopted by several workers in the field of shape recognition (see Attneave, & Arnoult, 1956, for a general review). Sets of stimuli are constructed according to certain probabilistic rules, and the main interest of the experimenter is in the comparison of average performance results for sets of figures generated by different rules, or variations of a given rule, rather than in the results obtained by comparing individual figures.

It should be pointed out at once that while this procedure makes it possible to formulate statements about properties of the stimuli presented to the subjects, this need not mean that the subject is necessarily responding to those properties—utilizing advantages theoretically present in the material, or being deterred by difficulties deliberately introduced. But, the use of stimuli of this kind does allow one to make predictions on an a priori rather than an ad hoc basis.

In practice, the rules employed generally permit the generation of large numbers of figures, so that one nearly always operates on samples drawn from given populations. Thus, a basic assumption of the method is that representative samples of figures drawn from the same population provide statistically identical inputs, and that differences among sample means, and within the samples themselves, can be regarded as random. Nevertheless, the probabilistic approach does not preclude interest in individual figures: it may, for instance, be of interest to establish which phenomenological properties resulting from the application of a certain rule make for ease or difficulty of discrimination, and whether a certain figure is always easy to identify, or only when it is presented in a given sample.

The aims of workers using this approach may differ widely. Reported herein are some of the results of a research program aimed at identifying factors making for ease or difficulty in discriminating amplitude-modulated signals displayed on a cathode ray tube.[1] One way of tackling the problem would have been to record a variety of existing signals, and to search for common factors in easy and difficult patterns. The investigators chose the alternative of imposing structure from without by generating a simulated set of signals on a probability basis. It was felt that this approach was likely to throw light on general problems of visual perception. The theoretical background and details of some of the methods used have been described in a paper by Fitts, Weinstein, Rappaport, Anderson & Leonard (1956).

[1]This research, conducted at the Laboratory of Aviation Psychology of The Ohio State University, was supported under Contract Nonr-495(02) between the Office of Naval Research and the Ohio State University Research Foundation.

The Stimuli

The figures employed in these particular studies are called Metric Figures. They consist of small, black-on-white histograms generated in two-dimensional matrices (see Fig. 1). In most of the studies the stimuli were presented on cards 1.6 by 1.6 inches in size, set in a plastic cover. The usual procedure was first to draw out the figures by hand, and then reproduce them by a photographic process.

Statistical Properties

1. Complexity. In a 4 x 4 matrix it is possible to generate 4^4 or 256 different solid figures, while in an 8 x 8 matrix the number of possible figures rises to 16,777,216. In information-theory terminology, random selection of one figure from each of these two populations can be said to generate 8 or 24 bits of information respectively. The basic dimension of complexity can thus be defined for this material in terms of the number of different figures which can be generated from a given matrix, and one would expect operations performed with the larger sample of figures to be more difficult than those performed with the smaller.

2. Redundancy. The population sizes cited above are the maximum possible when, in each figure, the column heights are selected independently of each other. As soon as any constraint is placed on this selection process, the number of figures which can then be generated will be smaller. If, for example, it is decided that in each figure each column height shall be present only once, then the total population of possible figures is reduced from n^n to n! (in the case of the 4 x 4 example, from 256 to 24). Such constraints introduce a dorm of redundancy, expressed in information-theory language as one minus the ratio of actual to possible information. Another way of introducing redundancy is to make one part of the contour dependent on another, as is done in mirror-imaging, or in repeating the figure. Examples of "random" and "constrained" figures are given in Fig. 2.

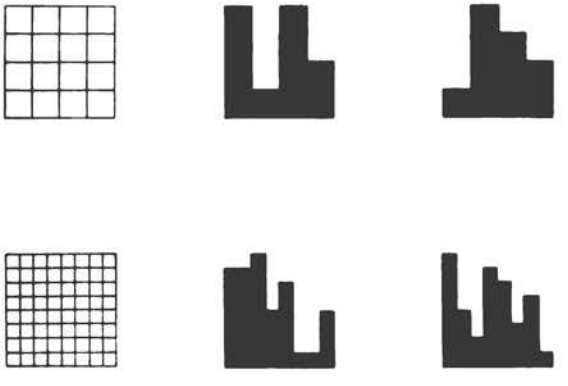


Fig. 1.--Two Sizes of Cell Matrices, With Samples of Metric Figures

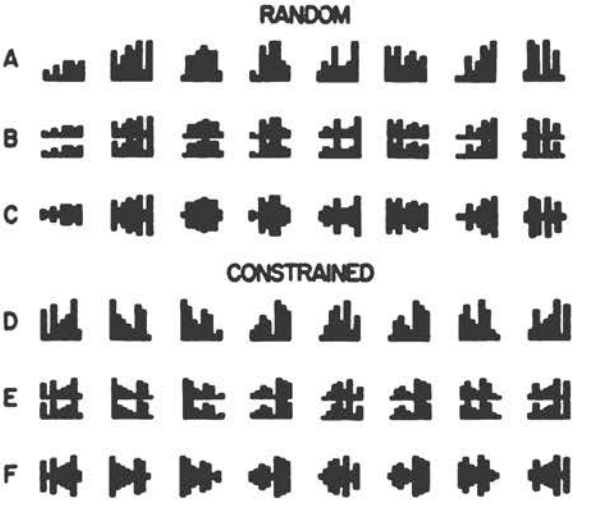


Fig. 2.--Examples of "Random" and "Constrained" Metric Figures.

Note:--The Figures in Rows A, B, and C were generated by randomized sampling. Those in Rows D, E, and F were generated by sampling without replacement.

3. Noise. Yet another form of statistical operation which can be performed is to subject each figure to specified amounts and kinds of distortion or "visual noise." Just as one generates the figures by a probabilistic process of selecting successive column heights, so is it possible to distort the contour of a figure by adding or subtracting incremental units from various columns according to a chosen set of probabilities.

In addition to varying the amount of "noise," one can also vary its type. One way of doing this is to change the shape of the probability function according to which the distortion is introduced, e.g., by making it more probable that columns will be increased than that they will be decreased in size. A second method is to make different parts of the figure available for distortion, e.g., rendering any cell in the matrix, rather than only those on either side of the contour, liable to a change of state.

Physical Properties

Apart from these statistical operations one can also manipulate physical properties of the stimuli, such as the orientation of the base line of the figures, or the size of their details.

Tasks Used

Once the principles defining the stimulus material have been specified, it is necessary to formulate various tasks on which to measure the effects of changing certain variables.

The Standard Identification Task

After a sample of figures had been generated, seven copies of each figure were obtained. The experimenter retained one set, and each of the other six sets was laid out in a row on a 36 x 40-inch board, as illustrated in Fig. 3. The board was placed slanting upwards at an angle of 10° with its lower edge 40 inches from the floor. Each row of figures was laid out in a different order, with equal spacing between the figures. The subject stood in front of the board, the experimenter next to him. When the subject was ready, the experimenter placed the first figure of his own set in a space above the uppermost row of figures on the sorting board and allowed the subject to look at it for two seconds. At the end of this period, a shutter was released which up to then had covered the six rows of stimuli, and the subject started searching for the duplicate of the test figure in the first row. Having found it, he flipped it over and continued successively through the remaining rows of the board, pressing a button to stop the timer after the test figure had been successfully identified in the sixth row. The turned-over figures were then replaced, and the procedure was repeated until all figures in the sample had been tested. The experimenter then either went on to a different sample, or, reshuffling the first, repeated the trial. With this procedure errors were rare and usually were not scored, the correction method being used during the identification procedure.

Other Tasks

In addition to the above, the material has been used in a simple sorting task, in a paced identification task, for paired-associate learning, and in a task requiring that each stimulus be reconstructed on a physical model. For most of the tasks

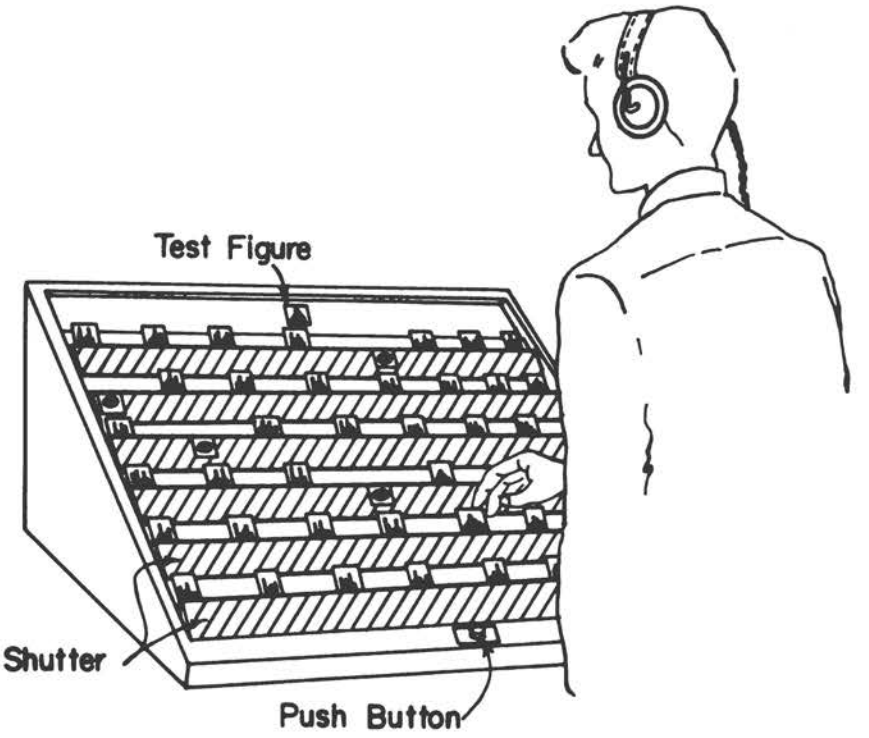


Fig. 3.--The Standard Sorting Board Used in the Speed of Identification Task.
Note:--Test figure appears once in each row. As soon as shutter opens, S locates test figures, turns them over, and pushes button to signify completion of a trial.

the stimuli were used in the form previously described, but for one of the paired-associate learning tasks the figures were made into slides and projected onto a screen for group testing.

Some of the Results

In the following account references are not noted in the text, but a list of papers, classified under the topics discussed, is provided at the end of the report.

Ease and Difficulty of Discrimination

As previously mentioned, the primary purpose of the research was to establish the conditions making for ease and difficulty of discrimination. An expected finding for the standard identification task was that the more complex a figure and the larger the size of the sample studied, the longer the time required to make a correct identification. In some cases mirror-imaging of the figures was found to have a beneficial effect, particularly after the subjects had had some practice. In general, vertical orientation of the base line of the figures resulted in greater ease of discrimination than horizontal orientation. Although most of the above effects were relatively small (of an order of about 10 per cent), they were usually sufficiently consistent to produce statistical significance.

The largest (and somewhat unexpected) effect was observed when those figures whose column heights had been selected independently of each other were compared with figures in which only one column of each height was allowed to be present. In a series of studies, it was found repeatedly that about 30 to 40 per cent more time was required for identification of the latter ("constrained") figures than for the former ("random") ones.

Another way of making discrimination more difficult is to distort the figures. Although fewer experiments have been done in this area, a number of findings warrant mention.

First, it can be noted that the methods of distortion used were effective in producing difficulties in discrimination: the greater the amount of statistically defined distortion imposed on a set of figures, the more difficult they were to identify. Part of the increased difficulty of identification appeared to be attributable to changes in similarity relationships produced among the members of a sample of figures by various amounts of distortion. (Parallel results have been obtained in studies of speech recognition at different levels of noise.) Yet another finding was that stimulus identification under paced conditions was as efficient when the subjects were given a relatively long look (three seconds) at a single distortion of a figure as when three different distortions of that figure were presented in the same length of time. Lastly, it was found that the effect of distortion could be reduced to some extent by introducing redundancy in the form of mirror images.

Follow-up of Certain Effects

Since Metric Figures are usually generated in n x n matrices, a variation of "complexity" can entail changes both in the number of columns and in the number of possible column heights. To determine which of these factors contributes more to increased difficulty of discrimination, a series of studies was carried out using samples of figures in which the number of columns was held constant, while the number of column heights was varied, and vice versa. Analysis of the results showed that increases in the number of columns were more detrimental to discrimination than increases in the number of column heights.

The finding that the "constrained" figures were harder to identify was somewhat surprising, since according to information theory such figures are redundant with respect to "random" ones. However, introducing redundancy by restricting the number of possible column heights renders a sample of stimuli more homogeneous from the phenomenological standpoint: the resulting figures are all equal with respect to area, have the same number of contour breaks, etc. Since the "constrained" figures look more similar to each other, it was hypothesized that phenomenological factors may have been responsible for the obtained results.

To test this hypothesis, additional stimulus samples were generated at the same level of complexity but under different rules of constraint, and the effects were studied for a number of different tasks. It was found that only for a task in which the subjects had to deal with each detail of a figure by reconstructing it on a physical model was there no difference in performance due to the rules of sampling. For all other tasks, which involved mere identification or naming, the original effect was found to hold, and degrees of constraint intermediate between "random" and "complete" were shown to produce intermediate effects on discrimination. It was, therefore, tentatively concluded that the introduction of constraint reduces the probability that peculiar identifying contours will appear in a given sample of fig-

ures, and that the observed increases in difficulty of identification are attributable to this effect.

Identification and Learning

One of the tasks employed in the research involved paired-associate learning: the subjects were required to learn the correct letter for each of 12 figures. Stimulus samples generated under three different rules were studied using different groups of subjects. It was found that those subjects who had to learn letters for a sample of "random" figures obtained the highest scores, while those working with the "constraint" sample scored lowest. The effects were of the same order of magnitude as those obtained for the two identification tasks utilized.

These results point up the role played by stimulus discriminability in the learning of lists of items. A point of general methodological interest illustrated by the study is that the probabilistic method of stimulus generation permits one to produce lists of quasi-nonsense material of differential difficulty.

Perceptual Learning

When a subject sorts through the same sample of figures for several trials, he shows a fairly continuous improvement in performance. It would be of interest to know how much of this improvement is due to increased familiarity with the whole class of figures, the particular population from which the sample was drawn, the members of the sample itself, or the specific nature of the task, etc. A start on this problem has been made by comparing the performance of subjects who work on the same sample of figures throughout a series of trials with that of subjects who change halfway through the experiment from one stimulus sample to another of the same level of complexity.

As might be expected, it was found that the group which changed stimulus samples showed considerable positive transfer. Subsequent experiments employing variants of the control condition indicated that only part of this transfer could be ascribed to perceptual learning of the particular class of material: the rest was found to be due to learning of the perceptual and motor aspects of the task itself.

Some Further Uses of the Method

Distorted Metric Figures can be used for a number of experimental purposes other than those described here. For example, the method can be employed in studies of concept formation to present subjects with stimulus material differing by known amounts and in specifiable ways from a standard figure to be detected or constructed. The task could comprise either the relatively simple problem of finding and describing the prototype figure, or the more complex one of identifying the rules used in generating the distortions.

Another possible use of the method is in the simulation of visual inspection tasks. For such experiments, known, suspected, or desired probability distributions of "faults" could be generated on specified kinds of visual material.

Concluding Remarks

In the preceding sections, a brief account has been given of an American military research program with which the author was associated for two years. It is

still too early to evaluate the extent to which the techniques described will contribute to basic understanding of human perceptual processes. At present, however, it can be said that probabilistic stimulus generation constitutes a useful research tool by means of which visual "nonsense" material can be graded on an a priori basis. In addition, it facilitates study of the effects produced on perception by objectively specified changes in the defining attributes of certain kinds of visual stimuli. Thus, the results reported here indicate that stimulus material of the type studied is best discriminated when the figures employed are of low complexity, have column heights selected on a "random" basis, and are mirror-imaged about a vertical axis.

References

General:

Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. Psychol. Bull., 1956, 53, 452-471.

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. J. exp. Psychol., 1956, 51, 1-11.

Complexity:

Weinstein, M. Stimulus complexity and the recognition of visual patterns. Unpublished doctoral thesis, Ohio State Univer., 1955.

Redundancy:

Rappaport, M. The role of redundancy in the discrimination of visual forms. J. exp. Psychol., 1957, 53, 3-10.

"Random" vs. "constrained":

Anderson, N. A., & Leonard, J. A. The recognition, naming and reconstruction of visual figures as a function of contour redundancy. (In preparation)

Distortion:

Rappaport, M. The role of redundancy in the discrimination of visual forms. J. exp. Psychol., 1957, 53, 3-10.

Leonard, J. A., & Fitts, P. M. Factors affecting the identification of distorted visual patterns. (In preparation)

Perceptual learning:

Leonard, J. A., Clarke, H. W., & Stoats, S. R. An experimental investigation of the factors contributing to general vs. specific perceptual learning. J. exp. Psychol., 1957, 53, 324-329.

Summary

The paper describes a military research program in which stimulus material was generated on a probabilistic basis for use in studies of visual perception and learning. The main purpose of the research was to identify stimulus attributes contributing to ease and difficulty of visual shape perception by human subjects.

Probabilistic stimulus generation permits an experimenter to obtain stimulus samples which differ from each other by known amounts along specified dimensions. It therefore permits more precise study of the perceptual effects produced by varying selected defining attributes of the sets of stimuli studied.

Since the rules of generation are usually such as to define large populations of stimuli, it is necessary in practice to work with samples from such populations. Thus, a basic assumption of the method is that mean performance differences among stimulus samples drawn from the same population follow a chance distribution. It is apparent that emphasis in the method is on the comparison of average performance results for sets of stimuli generated according to different rules. Comparisons based on individual figures, though not precluded, are usually of lesser interest.

The present research employed "metric figures": small, black-on-white histograms generated in two-dimensional matrices (see Fig. 1). Application of appropriate statistical rules served to define stimulus populations differing from each other with respect to complexity, redundancy, and distortion.

"Complexity" was defined in terms of the number of possible columns and column heights of the matrices from which the figures were generated: the 8^8 different figures which can be constructed from an 8 x 8 matrix are larger in number and more "complex" than, for example, the 4^4 possible figures which can be generated from a matrix of size 4 x 4.

"Redundancy," defined as one minus the ratio of actual to possible information in a sample of stimuli, was varied by: (a) placing specified constraints on the variation of histogram column heights (Fig. 2), and (b) making one part of a figure's contour dependent on another (through mirror-imaging or repetition of the figure).

"Distortion," or "visual noise," was introduced by: (a) altering various column heights according to a chosen probabilistic function, and (b) changing the shape of the probability function selected. Operations of type a vary the amount, and operations of type b alter the kind of distortion produced in a stimulus sample.

Also studied in the experiments were the effects on performance produced by manipulating such physical properties of the figures as orientation of the base line.

The tasks employed included: (a) figure identification under paced and unpaced conditions, (b) simple sorting of the figures, (c) paired-associate learning, and (d) a task requiring physical reconstruction of the figures. In the figure-identification task (Fig. 3) the subjects were asked to search a sample of stimuli for the duplicates of various test figures; in the learning task, they had to associate preassigned letters with the constituent figures of a sample.

Among the findings are the following:

1. Complexity is negatively related, and redundancy produced by mirror-imaging of the figures is positively related to stimulus discriminability. Ease of discrimination is facilitated by vertical (as opposed to horizontal) orientation of the histogram base lines. These effects, measured in terms of the time required to identify duplicates of the test figures, were small but usually statistically significant.

2. Difficulty of identification increases with the degree of statistically defined distortion imposed on a stimulus sample. This effect is apparently due to changes in inter-stimulus similarity relationships produced by "visual noise," and can be diminished by introducing redundancy in the form of mirror images. Discrimination was found to be equally efficient when the subjects were: (a) allowed to view a single distortion of a test figure for three seconds, and (b) given one-second glimpses of three different distortions of the same figure.

3. Decreasing the number of possible column heights was found to increase markedly the difficulty of discrimination for figures equal with respect to number of columns. This result was unexpected, since, according to information theory, constraints of the kind imposed serve to increase redundancy. Since the effect was found to hold for tasks involving identification, but not for a task requiring physical reconstruction of the test figures, the finding was tentatively attributed to the greater phenomenological homogeneity produced in a sample of stimuli by the introduction of constraint.

Analysis of performance on tasks involving learning showed that subjects presented with "constrained" stimulus samples (see above) had more difficulty learning to associate pre-assigned letters with the individual figures than subjects presented with stimuli whole column heights were allowed to vary randomly. These results, congruent with those obtained under the same conditions for the identification task, point up: (a) the role played by stimulus discriminability in paired-associated learning, and (b) the fact that probabilistic methods can be used in learning experiments to generate quasi-nonsense stimulus material of variable difficulty. It was also found in the learning experiments that the positive transfer manifested by subjects who changed from one stimulus sample to another halfway through a session was attributable partly to perceptual learning of the particular class of stimulus material employed, and partly to learning of the perceptual and motor aspects of the task itself.

Probabilistic methods of stimulus generation can also be employed, e.g., in research requiring the stimulation of visual inspection tasks, and in experiments on concept formation. By way of summary, it is emphasized that the method can be used: (a) to grade visual nonsense material on an a priori basis, and (b) to facilitate more precise determination of the effects produced on perception by objectively specified changes in the defining attributes of certain kinds of visual stimuli.

Résumé

L'article expose un programme de recherche militaire dans lequel des objets-stimulus ont été engendrés sur une base de probabilités pour être utilises dans des études de perception visuelle et d'apprentissage. Le but principal de la recherche consistait à identifier les propriétés des stimuli qui contribuent à déterminer la facilité et la difficulté de la perception visuelle des formes par les sujets humains.

La production de stimuli d'après des probabilités permet à l'expérimentateur d'en obtenir des échantillons qui diffèrent les uns des autres de façon connue selon des dimensions specifiees. Il permet donc une étude plus précise des effets de perception produits en variant certaines caractéristiques des groupes de stimuli étudiés.

Comme les règles de production sont généralement de nature à définir des populations de stimuli très grandes, il faut dans la pratique utiliser des échantillons pris dans ces populations. La méthode donc suppose que les différences moyennes de perception observées pour des échantillons de stimuli pris dans la même population suivent une répartition au hasard. Il est évident que la méthode met l'accent sur la comparaison des résultats moyens obtenus pour des ensembles de stimuli engendrés suivant diverses règles. La comparaison de stimuli individuels, bien que permise, a généralement un intérêt moindre.

La présente recherche a utilisé comme stimuli des "figures métriques": des petits histogrammes en noir et blanc engendrés dans des matrices à deux dimensions (voir Fig. 1). L'application de règles statistiques appropriées a servi à définir des populations de stimuli différant en matière de complexité, de redondance et de distorsion.

La "complexité" a été définie d'après le nombre de colonnes et de hauteurs de colonnes possibles des matrices dont ont été engendrées les figures; les diverses figures au nombre de 8^8 qui peuvent être obtenues d'une matrice de grandeur 8 x 8 sont plus nombreuses et plus "complexes" que, par exemple, les figures au nombre de 4^4 qui peuvent être engendrées d'une matrice de dimensions 4 x 4.

La "redondance," définie comme l'unité moins le rapport "information existante: information possible" dans un échantillon de stimuli, a été variée (a) en restreignant la variation possible des hauteurs des colonnes des histogrammes (Fig. 2), et (b) en faisant dépendre une partie du contour d'une figure d'une autre partie (en réfletant ou en répétant la figure).

La "distorsion" (ou "bruit visuel") a été introduite: (a) en changeant les hauteurs des colonnes d'après une fonction de probabilité choisie, et (b) en changeant la fonction de probabilité choisie. Des opérations du type a modifient la quantité, et des opérations du type b modifient le genre de distorsion produite dans un échantillon de stimuli.

Au cours des expériences, on a également étudié les effets sur la perception produits en manipulant des propriétés physiques des figures, telles que l'orientation de la ligne de base.

Les tâches expérimentales comprenaient: (a) l'identification des figures dans des conditions temporelles limitées et non limitées, (b) le simple triage des fig-

ures, (c) l'apprentissage des lettres associées préalablement aux figures, et (d) une tâche exigeant la reconstitution physique des figures. Dans la tâche d'identification, on demandait aux sujets de scruter un échantillon de stimuli aux fins de trouver le double d'une figure donnée (voir Fig. 3).

Parmi les constatations sont les suivantes:

1. La complexité est en relation inverse, et la redondance produite par le reflet des figures est en relation directe, avec la facilité d'identification des objets-stimulus. La distinction des stimuli a été facilitée par l'orientation verticale (par opposition à l'orientation horizontale) des lignes de base des histogrammes. Ces effets, mesurés en fonction du temps nécessaire pour identifier les doubles des figures d'essai, étaient petits mais généralement significatifs au point de vue statistique.

2. La difficulté d'identification augmente avec le degré de distorsion statistique imposée à un échantillon de stimuli. Cet effet est apparemment dû à des changements dans les rapports de similitude entre les figures produits par le "bruit visual," et peut être diminué en introduisant de la redondance sous forme de reflets. On a constaté que la distinction était tout aussi efficace quant les sujets pouvaient: (a) regarder une seule déformation d'une figure pendant trois secondes, et (b) apercevoir trois déformations différentes de cette figure pendant la même durée.

3. La diminution du nombre possible de hauteurs de colonnes a rendu beaucoup plus difficile la distinction de figures égales en nombre de colonnee. Ce résultat était inattendu: d'après la théorie d'information, des contraintes de l'espèce imposée servent à augmenter la redondance. Comme on a observé cet effet pour la tâche d'identification mais pas pour la tâche exigeant la reconstitution physique des figures, on impute la constatation à la plus grande homogénéité phénoménologique produite dans un échantillon de stimuli par l'introduction de contrainte.

4. Les expériences d'apprentissage ont montré que les sujets auxquels on avait présenté des stimuli restreints comme ci-dessus, éprouvaient plus de difficultés à y associer des lettres que les sujets confrontés avec des stimuli dont les hauteurs de colonne pouvaient varier au hasard. Ces résultats, qui concordent avec ceux obtenus dans les mêmes conditions pour la tâche d'identification, soulignent: (a) le rôle joué par la différenciation de stimuli dans l'apprentissage du type étudié, et (b) le fait que des méthodes statistiques d'engendrement de stimuli peuvent être utilisées dans des expériences d'apprentissage pour produire de la matière quasi-absurde de difficulté variable. On a aussi constaté au cours de ces expériences que le transfert positif manifesté par des sujets qui passaient d'un échantillon de stimuli à un autre au mileau d'une séance pouvait être attribué en partie à l'apprentissage des propriétés des stimuli employés, et en partie au fait d'avoir appris les demandes de la tâche elle-même.

La méthode statistique de la production de stimuli peut également être employée dans, par exemple, des recherches exigeant la simulation des tâches d'inspection visuelle, et des expériences sur la formation des concepts. En guise de résumé, on souligne que la méthode peut être utilisée: (a) pour estimer la difficulté des stimuli expérimentaux sur une base a priori, et (b) pour faciliter la détermination plus précise des effets produits sur la perception par des changements spécifiés objectivement dans les caractéristiques de certaines espèces de stimuli visuels.

594869 O - 61 - 15

Commentary by Discussant

Dr. Ivo Kohler (Austria):

Information theory, which looks at perception from the standpoint of transmitting messages to the subject from the environment, has thrown new light on many hitherto hidden aspects of perceptual processes. Research carried out within the framework of the theory has uncovered many new and interesting facts. What impressed me most in reading Dr. Leonard's paper was his report that, in one study, difficulty of discrimination was found to increase with the degree of redundancy introduced into the experimental stimuli. In order to understand this finding, it seems to me that two independent stimulus dimensions must be posited—one running from "simplicity" to "complexity," and the other running from "randomness" to "order." Thus, one might have "random" stimuli which are nevertheless simple, and "ordered" stimuli which are highly complex. I believe it would be of great interest to attempt to vary these two dimensions independently and investigate stimuli having all four possible combinations of high and low standings with respect to them.

Another way of approaching this finding is to speculate concerning its implications for adaptation. From the biological standpoint, an organism adjusting to its environment is faced with two main tasks. One is to perceive regularities in happenings around itself, and the other is to recognize exceptions to these rules. For the first task, redundant processes in the environment are valuable and even necessary, since, in a world of the purely random, neither learning nor adjustment would be possible. The second task, however, requires that the organism be sensitive to everything new and unfamiliar, so that it can successfully protect itself from catastrophic accidents. It may be, therefore, that less redundant stimuli are sometimes easier to recognize, because they possess certain novel characteristics to which the subject is tuned to respond.

Reply

Dr. Paul M. Fitts (U. S. A.):

Since Dr. Leonard was unable to attend the Symposium, I shall have to take responsibility for replying to Dr. Kohler's remarks. This will be easy to do since I find myself agreeing very heartily with what he has said. It is quite obvious that the research in question was markedly influenced by information theory, which was used in an attempt to define the experimental stimuli more precisely. Using the laws of redundancy to restrict the generation of the stimulus figures produced some unexpected results, and I was very much interested by Dr. Kohler's remarks on the theoretical implications of the findings.

THE DEVELOPMENT OF TRAINING CRITERIA WITH A VIEW TO OPERATIONAL VALIDITY

F. H. Lakin
Army Operational Research Group
Surrey, England

During World War II, and even to a great extent since, the only criteria of proficiency readily and commonly available to psychologists working with the British Army were those provided during training. The purpose of this paper is to consider how and to what extent performance measures obtained during the training stage of a man's army career can be made to serve as predictors of future proficiency on the job.

Training for the Job

All army training, whatever its nature, aims at producing men who will perform efficiently in their jobs. At one or more points during the period of instruction, the trainee is given some kind of trade test or examination, or is graded as to proficiency by his instructors. It is generally assumed that such tests, examinations, and grades are reliable, in the sense that they would yield the same result if readministered under comparable conditions to the same sample, and valid in the sense that they serve to assess or predict performance on the job.

The relations between the training program, the tests administered during training, and the actual content of the job are obviously of vital importance. Unless there is a high correlation between performance on the training tests and proficiency on the job, any given selection instrument may be valid in relation to the tests but not in relation to performance in the field, or vice versa. In the first case, selection based on instruments valid for predicting success in training may result in the elimination of men who are actually fitted for the job, or in the acceptance of men who are, in fact, unable to perform it. In the second case, selection based on tests valid for predicting job success may lead to irrelevant failures in training.

In using tests administered during the training stage as criteria of job proficiency there are two possible sources of error: (a) the training program may not be sufficiently related to job requirements, and (b) the trade test, final examination, or other index of proficiency may not be a reliable and valid measure of the knowledge or skill acquired during training.

Criteria of Job Proficiency

The development of on-the-job proficiency measures is the basic problem which must first be considered. Such measures, when practicable, provide yardsticks in terms of which one may evaluate.

1. Changes in an individual's performance on the job;
2. The relative proficiency of different individuals;
3. The effects of changes in training;
4. The results of changes in selection methods;
5. The validity of predictions made during selection and training;
6. The effects of modifications in equipment design.

In the case of a selection or trade test given during training, validity can be measured operationally in terms of the extent to which the test predicts performance on the job, but, in the case of job proficiency measures, one must be certain that what is being measured is of direct value in its own right and capable of being reliably assessed. In most army jobs there is no "unit of output" and no one dimension in terms of which to evaluate performance.

Two ways of approaching the assessment of job proficiency can be distinguished. The first is an analytic one: the job is studied in great detail and broken down into meaningful subtasks for the purpose of identifying critically important dimensions for objective measurement. In the second, more global approach, overall proficiency evaluations are obtained from those under whom the worker carries out his duties.

The two approaches can, of course, be combined in any particular study. The first is obviously the more objective, and for jobs containing reliably measurable elements of established importance for overall performance, it can be expected to yield results superior to those based on the opinions of supervisors.

Formidable difficulties, however, lie in wait for those who attempt the analytic approach. Among them are the problems of: (a) defining job elements in such a way as to be measurable yet psychologically meaningful, (b) selecting and combining measurable elements in such a way as to yield a "true" measure of overall performance, and (c) obtaining adequate measures of proficiency in the actual working situation. There is no known way, for example, of weighting speed and accuracy of rifle fire, marching, and speed of digging a foxhole so as to give an overall measure of a rifleman's combat performance, even supposing that all of these elements were reliably measurable in an operational situation. Under (c) one can cite the example of radar operators, whose ability to spot, select, and track targets on a Plan Position Indicator cannot be adequately measured under active service conditions, or even during exercises.

Describing and analyzing job performance in really meaningful terms is, thus, a difficult task, but one of great importance for test validation and the planning of training programs. For the purpose of this paper, job requirements are classified under the three general headings of: (a) practical skills, (b) theoretical knowledge, and (c) social skills and personal characteristics.

Practical Skills

Careful job study will often establish a variety of skills whose importance for total job performance is self-evident. The ability of a fitter to file to given limits, the ability of a radio mechanic to locate faults, and the accuracy of a rifleman in firing at a target, or the speed with which he can assemble or dismantle a given weapon, are all obvious factors in job proficiency.

Even in this domain, however, the number of job elements which can be satisfactorily measured in the working situation is disappointingly small. Adequate measures of such skills often can be obtained only by abstracting them from the operational situation. Unfortunately, even in the most realistic of simulated battle conditions, to cite an example, efficiency in many aspects of performance cannot be assumed to correspond exactly to efficiency in action, where reliable measures for the individual are nearly impossible to obtain.

Theoretical Knowledge

The extent to which theoretical knowledge contributes to job proficiency is arguable only in terms of some specific job or class of closely related jobs. If the job involves a variety of tasks unrelated to each other in terms of basic skills, then theory may be of little importance. On the other hand, some jobs involve a number of tasks consisting of variations on a basic theme, and resting on a common substratum of theory. To determine the importance of theory in job performance, it is necessary to determine how much theoretical knowledge is, in fact, retained by workers of differing levels of measured or rated proficiency.

Social Skills and Personal Characteristics

No military job can be analyzed purely in terms of technical skills and knowledge. Every soldier has to interact with others, often under conditions of intermittent stress or danger. Social behavior and reactions to stress cannot be measured in a complete or fully objective way, but it is possible to obtain data on how individuals differ in their reactions to critical situations. So far as social skills are concerned, since these are involved in the interactions of the individual with his superiors, equals and subordinates, ratings by others constitute more acceptable measures of "performance" in this area than they do in the domains of practical skills or knowledge.

Training Criteria

Measures of Skill and Knowledge

These can include marks in written tests and examinations, and objective measures of speed, accuracy, or level of performance in various job-relevant tasks.

Written tests administered during training can measure proficiency in certain aspects of job performance, provided that they yield reliable evaluations and cover clearly relevant areas. Map reading, arithmetic, radar theory, and procedures or sequences of operations, for example, constitute legitimate fields for such testing. It is important to insure that the measures developed for any particular job cover the work field as extensively as possible without including too great an amount of doubtfully relevant "theory."

A potentially more fruitful approach to the development of training criteria is the measurement of job aspects which can be abstracted from work situations without serious loss of realism, and objectively evaluated in terms of speed, accuracy, or some other suitable dimension. For example, in the operational situation it is very difficult, if not impossible, to obtain accurate measures of a radar operator's delay in detecting targets or ability to apply certain rules of target selection. However, the operator's task can be very closely simulated through the use

of artificial display devices. Using simulated radar apparatus makes it possible to measure delay and correctness of target detection under conditions where the nature and sequence of the signals appearing on the display panel can be exactly specified. Such synthetic training devices may take a variety of forms, ranging from virtually identical copies of apparatus actually used on the job to devices intended only to bring out the basic psychological and physiological factors involved in a given task.

If the test apparatus is identical with that used in the field (except for data-recording and control devices), then it can be viewed as a true work sample, and performance on it can be assumed to measure ability to carry out certain aspects of the job under slightly unreal circumstances. Even with near identity of simulation, however, some validation in terms of the relative performance of men of differing degrees of experience and rated proficiency is desirable.

It should also be emphasized that in simulated situations it is rarely if ever possible to reproduce all the factors which may affect performance in "real life". Differences with respect to such variables as stress, fatigue, danger, and level of motivation may render the test situation imperfectly valid as a measure of job performance in the field. The use of simulators is therefore limited by the extent to which their validity can be established.

Complex Skills, Social Abilities, and Personal Qualities

Finally, there remains a large area in which highly complex and as yet imperfectly defined factors must, for want of more objective means of evaluation, be rated, ranked, or in some way assessed by observers. In studies carried out in the United Kingdom on the selection of officers, the validity of assessments made during training for predicting performance after commissioning has rarely been high. For one thing, it is very difficult to obtain objective criteria of officer effectiveness, although the Critical Incident Technique provides a promising approach. In this procedure, descriptions of effective and ineffective behavior in specific situations are substituted for mere global opinions as to the individual's proficiency.

With respect to social skills in general, there is probably no more objective method of assessment than that based on the judgments of those who have taken part in the activities in which such skills were exercised. What others think to be social skill—or the lack of it—is probably as far as one can go toward an ultimate criterion. In this area, therefore, there is a strong case for using sociometric techniques through which superiors, peers, or subordinates rate or rank the subjects under consideration in certain defined ways. This approach has generally been far more popular in the United States than in the United Kingdom.

Conclusion

One of the most important and difficult areas of military psychology has been considered in this paper. In dealing with problems of selection, training, and equipment design, questions as to the soundness, and even the existence, of satisfactory criteria of proficiency are continually met. Evaluations of performance during training constitute an important source of criteria, partly because they are readily available and can be inexpensively obtained. Attempts to maximize the operational validity of training criteria, therefore, constitutes a task of great practical importance.

Summary

Progress in the areas of selection, training, and equipment design requires the development of satisfactory criteria of proficiency on the job. Evaluations of performance during training often constitute the only practicable or available method of estimating such proficiency. The present communication considers the problem of how the operational validity of assessments made during training can be maximized.

Using training measures to estimate future proficiency on the job involves two possible sources of error: (a) the training program may not be related to actual job requirements, and (b) the tests utilized may not be reliable and valid measures of the knowledge or skill acquired in training.

On-the-job proficiency criteria can be used to evaluate: (a) changes in an individual's job performance; (b) the relative proficiency of different individuals; (c) the effects of modifications in training methods, selection procedures, and equipment design; and (d) the accuracy of performance predictions made during selection and training.

Job proficiency criteria must yield reliable assessments and measure job elements of demonstrated importance to overall performance. Two ways of obtaining such criteria can be distinguished: (a) the analytic approach, which breaks the job down into subtasks for the purpose of identifying critically important and objectively measurable elements, and (b) the global approach, which relies on general proficiency evaluations based on the opinions of peers or supervisors.

The first approach is the more objective and can be expected to yield superior results when applied to jobs capable of meaningful analysis. Problems met by the user of this method include those of: (a) selecting and combining measured job elements in such a way as to yield satisfactory evaluations of overall performance, and (b) obtaining measures of proficiency in the actual work situation.

Job requirements can be classified under three general headings: (a) practical skills, (b) theoretical knowledge, and (c) social aptitudes and personality characteristics. Job analysis can establish a number of practical skills involved in proficiency, but these often cannot be measured without abstracting them from the total work situation in which they are imbedded. The importance of theoretical knowledge is sometimes overemphasized; to assess its contribution to performance in any particular job, one must determine how much theory is actually retained by workers of differing levels of proficiency. Social skills and personality characteristics are difficult to measure in a fully objective way, but criterion-relevant data can be obtained on how individuals differ, for example, in their reactions to critical situations. Ratings by others are more acceptable as measures of behavior in this area than they are in the domains of skill and knowledge.

As for training criteria, skill and knowledge can be assessed through written tests and performance measures. Use of simulated apparatus designed to reproduce as closely as possible the tasks involved in a job is an excellent means of estimating future performance. Such devices, since they permit precise control of many relevant variables, yield performance measures of a reliability not obtainable in the field. It is emphasized, however, that since in simulated situations it is rarely if ever possible to duplicate all the factors (such as stress, fatigue, danger, and

level of motivation) which affect operational performance, the predictive validity of performance tests is necessarily limited.

In the area of social skills and personality characteristics, assessments based on the judgments of subordinates, peers, and supervisors probably constitute the most objective means of evaluation available. The Critical Incident Technique and sociometric procedures hold promise as methods of measurement in this domain.

Since evaluations made during training are often the only measures of job performance which can be readily obtained, the development of operationally valid training criteria constitutes a problem of great practical importance.

Résumé

Pour progresser dans les domaines de la sélection, de l'entraînement et de la conception de l'équipement, il faut mettre au point des critères satisfaisants de la capacité dans le travail. L'évaluation du rendement pendant la période d'entraînement constitue souvent la seule méthode pratique ou disponible d'estimer cette capacité. La présente communication envisage des moyens de porter au maximum la valeur opérationnelle des estimations faites pendant la période d'entraînement.

L'utilisation de telles estimations pour prédire la compétence future dans le travail comporte deux sources possibles d'erreur: (a) le programme d'instruction pourrait ne pas correspondre aux exigences qui se poseront dans la pratique, et (b) les tests utilisés pourraient ne pas constituer des mesures sûres et valables des connaissances ou des habiletés acquises pendant l'entraînement.

Les critères de compétence dans la pratique peuvent être utilisées pour évaluer: (a) des changements dans le rendement du même individu, (b) la capacité relative d'individus différents, (c) les effets de modifications dans les méthodes d'instruction, les techniques de sélection, et la conception de l'équipement, et (d) l'exactitude des prévisions quant au rendement faites au cours de la sélection et l'entraînement.

Les critères de la capacité dans le travail doivent constituer des mesures sûres et évaluer des tâches partielles dont l'importance pour le rendement global a été eprouvée. On peut citer deux façons d'obtenir ces critères: (a) la méthode analytique, qui divise le travail en diverses tâches afin d'identifier des éléments critiques qui peuvent être objectivement mesurés, et (b) la méthode globale, qui s'appuie sur des évaluations générales de rendement faites par des surveillants.

La première méthode est la plus objective et on peut en attendre de meilleurs résultats lorsqu'elle est appliquée à des travaux qui peuvent être soumis à une analyse sensée. Des problèmes qui se posent à celui qui utilise cette méthode consistent: (a) à choisir et à combiner des éléments de travail de façon à permettre des évaluations satisfaisantes du rendement global, et (b) à obtenir des mesures de compétence dans la situation de travail réelle.

On peut classer les exigences de travail sous trois titres généraux: (a) les habiletés pratiques, (b) les connaissances théoriques, et (c) les aptitudes sociales et les caractéristiques de personnalité. L'analyse du travail permet de déterminer certaines habiletés pratiques comprises dans la compétence, mais celles-ci ne peuvent souvent être mesurées qu'en les détachant de la situation totale dont elles font partie. On accorde parfois trop d'importance aux connaissances théoriques; pour déterminer sa contribution au rendement dans un travail particulier, il faut établir la quantité de théorie que retiennent les travailleurs de divers niveaux de compétence. Les aptitudes sociales et les caractéristiques de personnalité sont difficiles à mesurer d'une façon tout à fait objective, mais on peut obtenir des données utiles concernant les façons différentes dont les individus réagissent dans des situations critiques par exemple. Des estimations faites par d'autres sont plus acceptables comme mesures du comportement dans ce domaine que dans ceux de l'habileté et des connaissances.

Examinant ensuite des critères disponibles pendant l'entraînement, l'habileté et les connaissances peuvent être évaluées par des tests écrits et des mesures de rendement. L'utilisation des appareils reproduisant étroitement les tâches que comporte un travail constitue un excellent moyen d'évaluation du rendement futur. Comme de tels dispositifs permettent un contrôle précis de nombreux éléments variables, ils donnent des mesures de rendement d'une sûreté qu'on ne peut pas atteindre dans la pratique. Il est souligné, cependant, que dans des situations simulées on peut rarement reproduire tous les facteurs (tels que la tension, la fatigue, le danger, et le niveau de motivation) qui affectent le rendement dans les conditions réelles. La valeur de prévision de tels tests est donc nécessairement limitée.

Dans le domaine des habiletés sociales et des caractéristiques de personnalité, des indications basées sur les jugements de subordonnés, d'égaux et de supérieurs constituent probablement le moyen d'évaluation le plus objectif dont on dispose. La Technique de l'Incident Critique et des méthodes sociométriques s'annoncent bien comme méthodes de mesure dans ce domaine.

Comme les évaluations faites pendant la période d'entraînement sont souvent les seules estimations du rendement dans le travail que l'on peut facilement obtenir, le perfectionnement de tels critères constitue un problème d'une grande importance pratique.

Commentary by Discussant

Dr. Robert Gagne (U. S. A.):

Mr. Lakin has given us a very good overview of the situation with respect to criteria and performance measurement, and, in so doing, has described a problem faced by all of us who are interested in the military applications of psychological knowledge. Although a great deal of attention has been given to this problem, the amount of actual progress made leaves much to be desired. In the U. S. Air Force, for example, it is still customary to validate selection tests against training cri-

teria such as course grades. However, the few studies using job performance measures have shown that the correlations of the tests with these criteria are noticeably different from their correlations with marks in training school. Thus, it may be that some significant improvements in selection would result from the development and use of performance measures as criteria.

As Mr. Lakin points out, moreover, such measures are needed not only as criteria against which selection tests can be validated but also as standards by which training procedures can be evaluated, and their goals specified. Only when we have satisfactory measures of job performance can we truly evaluate whether an individual has been adequately trained in the skills he is supposed to attain, and whether these skills are, in fact, required on the job.

Unfortunately, the difficulties in defining and measuring such behaviors are formidable. In the Laboratory I represent, which concerns itself with personnel responsible for maintaining complex electronic equipment, we are still trying to find objective records which will tell us how well a man performs various tasks on his job, (such as inspecting, adjusting, trouble-shooting, removing, replacing, and so on). Such records are very hard to find, and, even then, it is difficult to correct for all the sources of unwanted bias which they inevitably contain in their raw state. One alternative on which we have been working is the development of objective, job-performance checklists for use by supervisors.

Traditionally, our general approach to the problem treated by Mr. Lakin has been to define job performance criteria in terms of the specific activities demanded by the operational situation. The standards of performance for these activities can also be defined in terms of operational military requirements. In constructing situations which make the measurement of human performance possible, however, one is forced to abstract from the total situation in order to control unwanted variance. My own view is that the measurements derived from simulated situations of this kind can never be validated in the traditional sense. The main consideration is that they be acceptable to those who understand operational requirements.

Another characteristic of our approach has been our belief that the criteria for job-specialty training should be qualitatively the same as those for the job itself. In other words, it is held that training standards and job standards should differ only with respect to the level of performance, and should not differ in kind. To cite a specific example, if the job requires a man to adjust component X, then the training criterion should include a measure of his performance in adjusting component X. Thus, our tendency has clearly been to take what Mr. Lakin calls that more objective, analytic approach to the development of both job and training criteria. This general orientation might briefly be summed up by turning around the title of Mr. Lakin's paper to read: "The Development of Operational Criteria with a View to Training Validity."

General Discussion

Dr. Meredith P. Crawford (U. S. A.):

At the Army Human Resources Research Office, we have been giving considerable attention to the criterion problem. Two aspects of our research deserve mention. First, we have had great success in developing criteria when we have begun with job analyses carried out in operational or combat situations. For example, in a study of rifle marksmanship, we found from combat reports that infantrymen seldom engage targets beyond 300 yards, that they seldom actually see their targets, and that they usually have no time for windage adjustments. From observations like these, we were able to construct proficiency tests which required marksmen to fire at targets under conditions similar in many crucial respects to those encountered in actual battle. My first point, therefore, is that one can closely approximate operational criteria through careful field studies and interviews with operational personnel.

The second point I should like to emphasize is the importance of "face validity" in getting psychological research accepted by the military authorities. In other words, if we can develop proficiency tests of such obvious job relevance that military officials immediately recognize their usefulness as criteria, then we shall already have gone a long way toward gaining acceptance for new training procedures which result in proficiency on these criteria.

Dr. Clark L. Wilson (U. S. A.):

At Human Factors Research, Inc., we have made it a practice always to supplement our proficiency tests with ratings. This is done in an attempt to insure coverage of portions of the criterion variance which might not otherwise be measured. In one recent study we found a rather interesting thing: our predictors worked better for that quarter of a large group of subjects about whose performance our raters showed the least agreement. Lower test correlations were obtained for the quarter of the group about whose performance the raters agreed best.

These findings have yet to be cross-validated, but I thought they were interesting enough to deserve mention here. Usually, we assume that if two people agree in their assessment of a man with whose performance they are thoroughly familiar, then that judgment is likely to be correct. The results just described suggest that the opposite may be true. Perhaps if two raters see different aspects of a subject's performance, they may be covering that performance more thoroughly, so to speak, than two people who focus on the same behavioral dimensions and, therefore, tend to be more consistent in their judgments.

Reply

Mr. F. H. Lakin:

I should like to thank the discussants, and to say that when I first started writing this paper I felt that the United States probably had far more to contribute to this topic than we had. Nevertheless, I wanted to present our views on the problem and try to generate some discussion. I should like to add that I might very well have turned around the title and talked about operational criteria with a view to training validity. Actually, I was of two minds as to which way to do it—a fact which in itself exemplifies the central nature of the problem.

Contributors and Discussants

Bakan, Paul
Department of Psychology
Michigan State University
East Lansing, Michigan
U. S. A.

Baker, Lynn E.
U. S. Army Research and Development Office
Washington, D. C.
U. S. A.

Biesheuvel, Simon
Director, National Institute for Personnel Research
Johannesburg, Union of South Africa

Bonnardel, Raymond G.
Laboratory of Applied Psychology
The Sorbonne
Paris, France

de Brisson, A.
Air Force Center of Psychological Studies and Instruction
Versailles, France

Broadbent, Donald E.
Medical Research Council
Cambridge, England

Carlson, Col.
U. S. Air Force
U. S. A.

Carmichael, Leonard
Smithsonian Institution
Washington, D. C.
U. S. A.

Carp, Abraham
Air Force Personnel and Training Research Center
Lackland Air Force Base, Texas
U. S. A.

Chandessais, Lt. Col. Charles
Director of Human Sciences
Permanent Secretariat of the Scientific Action Committee of National Defense
Paris, France

Chapanis, Alphonse
Department of Psychology
The Johns Hopkins University
Baltimore, Maryland
U. S. A.

Christensen, Julian M.
Psychology Branch
Aero Medical Laboratory
Wright Air Development Center
Wright-Patterson Air Force Base, Ohio
U. S. A.

Coucheron-Jarl, Vidkunn
Psychological Division
Norwegian Armed Forces
Oslo, Norway

Crawford, Meredith P.
Human Resources Research Office
Department of the Army
Washington, D. C.
U. S. A.

Delys, Louis
Center of Psychotechnical Studies and Research
Belgium Armed Forces
Brussels, Belgium

de Dixmude, Baron P. Jaques
Committee Chairman
Belgian Armed Forces Chief of Staff
Brussels, Belgium

Edwards, Ward D.
Department of Psychology
University of Michigan
Anne Arbor, Michigan
U. S. A.

Fokkema, Sipke D.
Office of Selection Affairs
Royal Dutch Air Force
The Hague, The Netherlands

Gagne, Robert
Air Force Personnel and Training Research Center
Lackland Air Force Base, Texas
U. S. A.

Geldard, Frank A.
Chairman, NRC Committee on Military Psychology
Department of Psychology
University of Virginia
Charlottesville, Virginia
(Now Dean of Graduate School, U. of Va.)
U. S. A.

Gerhardt, Rolf
Psychological Division
Norwegian Armed Forces
Oslo, Norway

Germain, Jose
Department of Experimental Psychology
Superior Council for Scientific Investigation
Madrid, Spain

van der Giessen, R. W.
Royal Dutch Navy
The Hague, The Netherlands

Gualtierotti, T.
Laboratory of Physiology
University of Milan
Milan, Italy

Glaser, Robert
American Institute of Research
University of Pittsburgh
Pittsburgh, Pennsylvania
U. S. A.

Guttman, Louis
Institute of Applied Social Research
Jerusalem, Israel

Hamon, Joseph M.
Surgeon General, Military Health Service
Paris, France

Husen, Torsten
Institute of Educational Research
Teachers College, University of Stockholm
Stockholm, Sweden

Knoepfel, H.-K.
Air Force Psychiatric Service
Zurich, Switzerland

Köhler, Ivo
Institute for Experimental Psychology
University of Innsbruck
Innsbruck, Austria

Kris, Christine
United Kingdom

Kruisinga, R. J. H.
Leeuwarden
The Netherlands

Kryter, K. D.
Operational Applications Laboratory
Air Force Cambridge Research Center
Bolling Air Force Base, Washington, D. C.
U. S. A.

Lakin, F. H.
Army Operational Research Group
Surrey, England

Lange, Carl J.
Human Resources Research Organization
Carmel, California
U. S. A.

Legrand, P.
Air Force Center of Psychological Studies and Instruction
Versailles, France

Leonard, J. Alfred
Applied Psychology Research Unit
Medical Research Council
Cambridge, England

Mackworth, Norman H.
Applied Psychology Research Unit
Medical Research Council
Cambridge, England

Margeria, R.
Laboratory of Physiology
University of Milan
Milan, Italy

Melton, Arthur W.
U. S. Air Force Personnel and Training Research Center
Lackland Air Force Base, Texas
U. S. A.

Meschieri, Luigi
National Institute of Psychology
National Research Council
Rome, Italy

Mitchell, Col. Philip H.
Director, Human Factors Division
U. S. Air Force Air Research and Development Command
Washington, D. C.
U. S. A.

Myers, C. Roger
Department of Psychology
University of Toronto
Toronto, Canada

Penton, James C.
Army Operational Research Group
Surrey, England

Pinillos, J. L.
Department of Experimental Psychology
Superior Council of Scientific Investigation
Madrid, Spain

Rainio, R. Kullervo
Institute of Industrial Supervision
Helsinki, Finland

Riis, Erik
Psychological Division
Norwegian Armed Forces
Oslo, Norway

Sadler, P. J.
Air Ministry
London, England

Siro, Eljas
Institute of Psychology
University of Helsinki
Helsinki, Finland

Spinelli, D.
Laboratory of Physiology
University of Milan
Milan, Italy

Stockbridge, Hugh C. W.
Clothing and Stores Experimental Establishment
Ministry of Supply
London, England

Termöhlen, Major J.
Institute of Military Psychology
Copenhagen, Denmark

Ungricht, Jean
Academic Vocational Guidance Office
Zurich, Switzerland

Wallis, Donald
Department of the Senior Psychologist
The Admiralty
London, England

Webb, Wilse B.
Aviation Psychology Laboratory
U. S. Naval School of Aviation Medicine
Pensacola, Florida
U. S. A.

Wilson, Clark L.
Human Factors Research, Inc.
Whittier, California
U. S. A.

Wilson, Norman A. B.
Senior Psychologist
The Admiralty
London, England

Witte, Wilhelm
Institute of Psychology
University of Tübingen
Tübingen, Germany

Zubin, Joseph
New York State Department of Mental Hygiene
New York, New York
U. S. A.

U. S. GOVERNMENT PRINTING OFFICE : 1961 O - 594869