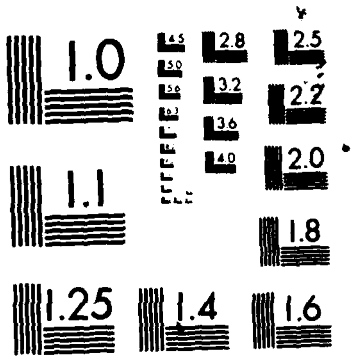
AD-A172 682 STANDARDIZATION OF PERFORMANCE TESTS: A PROPOSAL FOR FURTHER STEPS(U) TECHNISCHE HOCHSCHULE AACHEN (GERMANY F R) INST FOR PSYCHOLOG... A F SANDERS ET AL. 01 JUL 86 1/2

UNCLASSIFIED EOARD-TR-86-08 AFOSR-85-0305 F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

STANDARDIZATION OF PERFORMANCE TESTS:
A PROPOSAL FOR FURTHER STEPS

A.F. Sanders, R.C. Haygood, H.-W. Schroiff,
C.H. Wauschkuhn

Institute for Psychology, RWTH Aachen,
Federal Republic of Germany

Scientific Report, Phase I, 1 July 1986

Grant Number: AFOSR-85-0305

STANDARDIZATION OF PERFORMANCE TESTS: A PROPOSAL FOR FURTHER STEPS

A.F. Sanders (Principal Investigator),
R.C. Haygood,
H.-W. Schroiff,
C.H. Wauschkuhn

Institute for Psychology
Rheinisch-Westfälische Technische Hochschule Aachen
Aachen, Federal Republic of Germany

1 July 1986

Scientific Report, Phase I, 01 Sept. 1985 - 31 Aug. 1986

Prepared for

UNITED STATES AIR FORCE
Air Force Office of Scientific Research

and

EUROPEAN OFFICE OF AEROSPACE RESEARCH AND DEVELOPMENT
London, England

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: EOARD-TR-86-08	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER: AD-A172682
4. TITLE (and Subtitle): Standardization of Performance Tests: A Proposal for Further Steps	5. TYPE OF REPORT & PERIOD COVERED: Scientific Report 01 Sept. 85 - 31 Aug. 86
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): A.F. Sanders (principal investigator), R.C. Haygood, H.-W. Schroiff, C.H. Wauschkuhn	8. CONTRACT OR GRANT NUMBER(s): AFOSR-85-0305
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Institute for Psychology, Aachen University of Technology, Jägerstr. 17-19, D-5100 Aachen, West-Germany	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2301/D1 192
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE: 01 July 1986
	13. NUMBER OF PAGES: 179
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Human Performance; Task Batteries; Standardization

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes the results of Phase I of a study on the feasibility and development of a standardized task battery for measurement of human performance.
Phase I includes literature reviews on task batteries, interviews with experts in the field of human performance and analytic studies of the theoretical background of existing and possible task batteries.
Recommendations for a further task battery are given.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Standardization of Performance Tests: A Proposal for Further Steps		5. TYPE OF REPORT & PERIOD COVERED Scientific Report 01 Sept. 85 - 31 Aug. 86
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A.F. Sanders (principal investigator), R.C. Haygood, H.-W. Schroiff, C.H. Wauschkuhn		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Psychology, Aachen University of Technology, Jägerstr. 17-19, D-5100 Aachen, West-Germany		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 01 July 1986
		13. NUMBER OF PAGES 179
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Human Performance; Task Batteries; Standardization

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report summarizes the results of Phase I of a study on the feasibility and development of a standardized task battery for measurement of human performance.
Phase I includes literature reviews on task batteries, interviews with experts in the field of human performance and analytic studies of the theoretical background of existing and possible task batteries.
Recommendations for a further task battery are given.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR-85-0305

TABLE OF CONTENTS

1. INTRODUCTION

1.1. BACKGROUND OF THE PROJECT

A major problem today in human performance research is that researchers have used a variety of experimental methods and tasks. Even when the task is ostensibly the same (e.g., multiple-choice reaction time), experimenters have used different task parameters, equipment, stimuli, instructions, and so forth. This lack of standardization has created several problems for those who wish to use the results for practical decision making. For example there are no norms for the various experimental tasks. Furthermore, when there are differences in outcomes, they are often attributed to differences in method, without definitive evidence of what the relevant differences are. In fact, the documentation regarding procedures, equipment, subjects, and independent variables has frequntly been indadequate to the degree that exact replication of many experiments is impossible. Finally, there is a widespread complaint that the methods and tasks used in the laboratory are so simple and artificial that they have little or no applicability to real world tasks. Certainly there has been little attempt to relate laboratory tasks to real-life tasks or even to each other.

THE AACHEN MEETING

Dissatisfaction with this state of affairs during the early 1980's led to the scheduling of a meeting and workshop held in Aachen, Federal Republic of Germany, at the Institute for Psychology of the Rheinisch-Westfälische Technische Hochschule (RWTH) on 23 and 24 October, 1984. This meeting was sponsored and funded by the USAF European Office of Aerospace Research and Development (Grant SCP 85-1003). The meeting was attended by a broad spectrum of interested parties, including USAF, TNO (Netherlands), MRC, CERMA, etc.

At the Aachen meeting, the major topic of interest was the feasibility and desirability of development of a standardized battery of performance tasks for international use; a major emphasis for the battery was to evaluate the effects of environmental stress, including the effects of drugs, lack of sleep, prolonged excessive workload, etc. Such a battery was seen as having potential for use in both theoretical and applied research and in personnel selection. At a minimum, the use of a standardized version of each experimental task was seen as providing comparability of results across different research studies. The consensus of participants was that the development of a standardized battery was desirable and feasible, and that study of the problem should proceed as quickly as possible.

The results of the workshop indicated that there was general agreement regarding the desirability of including certain tasks (Sternberg memory search, tracking, continuous memory, and the Baddeley-Hicks task), and a variety of other popular candidates surfaced (e.g., perceptual encoding, sustained attention). General agreement was also obtained that each task included in the final battery should be supported by a definition of:

1. The theoretical basis of the task.
2. The corresponding aspects of real-life performance
3. Specific modes of operation -- equipment, task parameters, procedures, etc.
4. Norms for each relevant population.

The question of how best to proceed was discussed at length. In particular, concerns were voiced as to who should "lead" the effort, how it might be funded, the scheduling of future meetings, and so forth. In the end, the responsibility for leading the effort and securing funding was accepted by the RWTH Aachen Institute for Psychology and the current project was the result.

1.2. PROJECT DESCRIPTION

THE OVERALL PROJECT

The project was designed to take place in two phases, with a tentative third phase contingent on satisfactory results of the first two phases. A proposal to accomplish this work was submitted to the USAF European Office of Aerospace Research and Development on Feb. 5, 1985 and work began officially on September, 1, 1985. The following is an outline of the project:

Phase I - Literature Review, Interviews and Analytic Studies

1. Review of the literature on task batteries.
2. Selective Reviews of the theoretical literature on human performance tasks, as commonly found in task batteries.
3. Interviews with prominent persons in the field of human performance measurement and theory.
4. Integration of information and completion of detailed plan for Phase II. Submission of Phase I report.

Phase II. - Development and Laboratory Testing of Candidate Tasks

1. Selection of candidate tasks.
2. Programming and implementation of selected tasks on equipment at the Institute for Psychology, RWTH, Aachen.
3. Tryouts of tasks under both stressed and unstressed conditions, and revision of both battery content and individual task parameters and procedures.
4. Preparation and submission of Phase II final report; report will provide all detail necessary for implementation of the battery, and a detailed discussion of the human mental and physical functions represented in the battery, as well as relevant information concerning relevant information concerning the effects of stress on each task.
5. Preparation and submission of proposal for follow-on Phase III.

Phase III - Real-World Validation

The purpose of the third phase is to try out the battery tasks in various real-world settings, including both operational and simulator conditions. Both predictive and synthetic validation will be pursued, including an examination of the degree to which standardized battery performance can be used to predict success in training and in later job performance. The final output of this phase will be a preliminary cut at tying these laboratory tasks to performance in real world tasks.

PROJECT SCHEDULE

The original proposal envisioned that Phase I would require 18 months, and that Phase II would begin after 15 months and last for two years. Thus the total time for the first two phases would have been approximately 3 years, 3 months. It now appears that Phase I will require only 12 months; part of this improvement was achieved by beginning the interviews immediately instead of waiting for the completion of task battery reviews and analyses. Thus the first two phases should require about 3 years, and our current goal is Phase II completion in the Fall of 1988. The specific schedule contemplated is:

June 30, 1986	Submission of preliminary information concerning Phase I.
September 1, 1986	Submission of Phase I Final Report and plan for Phase II.
January 1, 1987	Phase II begins.
July 31, 1987	Phase II Progress Report.
July 31, 1988	Phase II Final Report.

PROJECT PERSONNEL

The following are brief descriptions of project personnel:

Principal Investigator: Andries F. Sanders, Ph. D.
Dr. Sanders is Professor and Director of the Institute for Psychology, RWTH, Aachen. He received his Ph. D. at the University of Utrecht, in the Netherlands. From 1957 until 1984 he was a scientist at the Institute for Perception, TNO, The Netherlands, where he rose to the positions of Head of the Experimental Psychology Department and Deputy Director of the Institute.
Aside from project administration, Dr. Sanders has designed and conducted some of the interviews and taken part in drafting the literature summaries.

Project Scientist: Hans-Willi Schroiff, Ph.D..
Dr. Schroiff is a senior staff member of the Institute for Psychology, conducting a variety of research into human performance, including the role of vision in driving performance. Dr. Schroiff received his Ph.D. at 1983 in Aachen.

He joined the staff of the project in literature reviews and conducting and reporting the interviews.

Project Scientist: Robert C. Haygood, Ph. D..
Dr. Haygood received his Ph. D. at the University of Utah in 1963, and has taught at Kansas State University and Arizona State University where he now holds the rank of Professor. He is serving as Guest Professor at the Institute for Psychology during the 1985-86 academic year. Dr. Haygood's major scientific interests are in adaptive training and in human performance measurement.
His contribution has been in reviewing the theoretical background of the performance measurement effort.

Project Scientist: C. Hilka Wauschkuhn, Diplom Psychologin.
Hilka Wauschkuhn received her diplom in Göttingen, FRG 1982. From 1983 to 1985 she has been coworker in a project on psycho-neuro-endocrinology at the Deutsches Primatenzentrum, Göttingen.
She joined the Aachen project in January 1986. She has primary responsibility for coordinating the efforts of other staff members and development of scientific documentation. Included in her responsibilities are that of performing analytic work regarding the interviews and reviews of scientific literature.

Some other members contributed to the project by summarizing some relevant topics in the area of human performance:

- Mike Donk, cand.-phil., received her Vordiplom at Tilburg/NL and is now doing the Hauptstudium at our insitute in Aachen, she wrote the chapter on time sharing and dual performance (5.4.2.).
- Will Spijker's contribution is the chapter on tracking performance (5.4.1.). Will Spijkers received his masters degree in 1978 from the University of Tilburg/ NL. Since that time he has been affiliated with the Insitute for Perception (TNO), and the Universities of Nijmegen and Tilburg, both teaching and doing research in human motor performance. He joined the staff of the Aachen institute in January 1985.
- Jan Theeuwes, cand.-phil., wrote the chapter on choice reaction processes (5.4.4.). Jan Theeuwes is doing his Hauptstudium of psychology in Aachen. He received his Vordiplom at the University of Tilburg/ NL.

1.3. ORGANIZATION OF THE REPORT

This report is organized according to the major tasks performed in Phase I, with a final section for conclusions and recommendations. To avoid overwhelming the reader, the bulky details of the interviews and the reviews of task-battery literature have been placed in appendices; a concise summary and discussion for each is given in the main body of the report. The following is a brief description of the major sections of the report.

a) Interviews--the first major effort of Phase I was conducting interviews on the feasibility of a standardized task battery with a number of prominent persons in the field of human performance research. A total of

25 interviews was conducted. The complete protocols or these interviews are provided in Appendix 5.2.; a discussion of the results and summary of the general trends in the opinions is found in Section 2.1.

b) Review of task-battery literature--this effort consisted of collecting information on the task batteries that are already in operational use or which are about to be completed. Information was obtained about seven batteries, five from the United States and two from Europe. Altough review of information about other batteries was anticipated, the necessary information did not arrive in time to be included in this report. However, we feel that the present set of batteries is generally representative of the kinds of batteries in use and in development. A full account of the task batteries reviewed is found in Appendix 5.3.. The results of our analyses and a summary statement of the main trends is found in Section 2.2.

c) General approach and theoretical considerations--it was necessary to consider in some depth both the elements of our approach to battery development and the theoretical backgrounds of potential candidate tasks. These are found in Section 3. A general review of the theoretical backgrounds underlying the most common tasks used in existing batteries was conducted. On the basis of the summary table of these tasks (see section 2.2.) it was decided to provide concise literature reviews on the topics of (1) manual tracking, (2) time sharing and dual performance, (3) visual processing, (4) perceptual-motorspeed and choice reaction processes, (5) memory search, and (6) lexical and semantic encoding. These reviews are reported full in Apendix 5.4.. A summary of some major concepts underlying task batteries--including the largely atheoretical factor analytic approach--is presented in Section 3.

d) Conclusions and recommendations--the main body of the report concludes with a section containing conclusions and recommendations, in which (1) the most popular tasks are briefly summarized, (2) some apparent gaps are discussed, (3) the major stands on background concepts are mentioned, and (4) some of the major issues about relating laboratory tasks to real life tasks are sketched. Finally some recomendations are formulated.

2. DATA BASE

2.1. INTERVIEWS ON THE FEASIBILITY OF A STANDARDIZED TASK BATTERY IN HUMAN PERFORMANCE RESEARCH

INTRODUCTION

This section summarizes the results of a number of interviews conducted with active researchers in the field of human performance during the fall of 1985. Many of the interviews were conducted by Dr. Schroiff during the "Conference of the Psychonomic Society 1985" (Boston, USA). Some interviews were conducted at the NATO-meeting in Les Arcs (France) by Dr. Sanders and Dr. Debus. Dr. Broadbent submitted his views in writing.

In the interviews the personal views of the interviewees towards a number of discussion topics were collected. The interviewees were briefed about the purpose and the contents of the research project by having them read a two-page outline of the project (see Appendix 5.1.).

The interviewees were asked the following questions:

(1) Which kinds of methods (experimental paradigms, performance-task settings) have you been using in human performance research?

(2) Which methods do you regard as particularly useful a) with respect to theoretical developments? b) with respect to generalizability to real life performance?

(3) Do you know about any metric except speed or accuracy that is useful in the assessment of skills?

(4) Could you comment on the reasons for the low validity of performance tests/ test batteries with respect to the prediction of performance in real life tasks?

(5) Do you have any ideas for improving the generalizability of such laboratory tasks?

(6) To what degree can a real life task be broken down into components that can be isolated and assessed separately?

(7) What do you think about the feasability of developing a standardized battery of performance tests? Which tests do you think should be included? What do you think about factor-analytic approaches?

(8) If interviewee is positive towards question 7) Do you think it is possible to develop a broad enough battery of tests to cover most of the important real life skills?

(9) Do you have any ideas on skill categories or classification of skills that should be considered in a project like this?

GENERAL REMARKS

It was pointed out that a standardized battery of laboratory tasks for human performance assessment could serve different purposes. First, the main aim could be directed at the assessment of differences between people. Second, the assessment of the effects of environmental variables could be the topic of interest. Finally, it could be of interest to assess the impact of some proposed new task on total performance. As Broadbent points out, the requirements for a battery would differ substantially depending on the purpose, so that in the end three batteries of tests might be needed instead of one. The three possibilities should be kept in mind during the further discussion of this project.

For further reading it seems necessary to differentiate on a conceptual level between "abilities" which are regarded relatively constant (e.g. visual acuity) and "skills" (e.g. visual search) which are subject to change by (e.g.) different strategies that are employed.

All interviewees were positive towards the general idea of the project. Everybody found it desirable to establish a standardized battery of tasks in order to achieve a better comparability between results from different laboratories, although it was felt that some people might not adopt a positive outcome of the project because they might feel themselves restricted in their "scientific creativity".

ANSWERS TO THE QUESTIONS

The tasks that are mainly employed in the domain of human performance research are: choice RT, tracking, STM/LTM tasks, dual task capacity, knowledge based skills (e.g. reading, arithmetic), tests of the knowledge base itself (e.g. reasoning, spatial ability), attention and vigilance tasks. The main measures reported by the interviewees are reaction time, physiological measures, and recall and recognition paradigms.

The following tasks should be included in a standardized task battery according to most of the interviewees:

- perceptual measures (e.g contrast sensitivity, visual acuity)
- STM-measures
- visual motor coordination
- speed of retrieving linguistic information
- Sternberg-tasks
- Tracking (stable, unstable)
- spatial information processing
- Embedded figures
- Dual-task tests (e.g. dichotic listening)

As will be pointed out below, the majority of interviewees, however, felt that the available laboratory tasks were not good candidates for the intended purpose because they were selected and developed for some other reason. Furthermore only tests or tasks should be selected that are predictive for the final performance level (i.e. after extended practice).

Everybody agreed that the battery should comprise not too many tasks. This, however, should depend on the degree of task-specific knowledge and complexity of the real-life task to be predicted.

There was a general agreement about the the low predictive validity of laboratory tasks with regard to real life performance. The main reason is probably that some extra function(s) or skill(s) that are (is) relevant in real-life performance will not be assessed in the laboratory situation. The opinions differ slightly with regard to the causes. Some people believe that this is due to the context-reduced nature of the laboratory task: most experimental paradigms are not aimed at evaluating all the variables that affect performance. On the contrary, they are designed to investigate a specific phenomenon that is artificially isolated by the experimental set-up.

Tests of isolated abilities or skills usually do not incorporate interaction effects when these skills have to be combined in a real-life task. Although the single components may be highly practiced this does not mean that the complex performance will be at the same high level. It is felt that until now there is no good way to assess the "assembly" of component abilties or skills. It is not surprising that (e.g.) laboratory tasks of visual search normally have a reasonably high predictive validity, because task parameters in laboratory and real life search do not change substantially. RT measures can only have a predictive value for real-life tasks if the subject in the real-life task is under comparable time constraints.

One generally finds a neglect of strategical aspects of behavior in laboratory research on human performance. Real-life performance seems to be more subject to strategical influences. Here again the artificial character of the laboratory experiment that seeks to deprive the subjects of their strategical freedom comes into play. One way to improve validity is to complement the traditional two-choice laboratory tasks with tasks with more performance alternatives. What obviously is needed are process models that to some extent dictate the meaning of performance measures. At the moment there are no good models available for such an analysis.

The level of practice also seems to be responsible for the low predictive validity. Compared to performance in real-life situations laboratory performance is usually little practiced. This means that the behavior has not yet reached its optimal level of organization and the integrating effects of extended practice have not worked out. Practice might change the underlying factorial structure of skills (see e.g. the results of Fleishman).

The problem seems to be best stated by a literal quote from Kahneman: "It is hopeless to believe that a preliminary test of a single skill should have predictive value for a highly practiced complex task where this skill interacts with numerous other skills and that interaction is directed by different strategical supervisors".

One should be careful, however, in attributing the low predictive validity solely to the factors mentioned above. Broadbent has argued that the low correlations between the results of aptitude tests for aviators and actual flight performance might simply stem from the low variability amongst the highly selected sample of persons who are admitted to flying training. Also a high degree of variability (i.e. poor reliability) of the prediction criteria may be one of the causes for a low degree of predictive validity.

The question of whether a break down of complex tasks into components is possible provoked a number of controversial statements. The general possibility of breaking down a not-too-complex task into its constituents was not denied, but the success of a venture like this is highly dependent on the quality of a task analysis. This should not be a task analysis in the classical sense but a cognitive component analysis. The general opinion is that this might work for a small number of well described tasks whose theoretical task structure and the hypothetical component processes involved are well known (e.g. car driving). Again it is argued that complex phenomena of human cognition cannot be broken down into a very few basic dimensions. Even if one would succeed here, the problem of assessing the interaction between the components remains. It is seen that the success of the research program will depend on the degree that a) basic conceptual units of human performance can be defined, b) adequate measurement procedures can be worked out to assess these basic skills or abilities, and c) a test can be devised that reveals the interindividual differences in the "assembly" of those skills and abilities in real-life tasks. It is felt that the more one decomposes, the less predictive validity can be expected.

This leads to a prominent alternative to a standardized battery: the use of simulation methods, which is regarded as the principal way to achieve a good prediction. The relative advantages and disadvantages of simulation should be worked out more clearly.

A second alternative seems to be the use of process models of task performance - a probably forthcoming research strategy in connection with the aims of this project. However, as pointed out above, this domain has been explored to a minor extent only.

The question with regard to the feasibility of a standardized task battery has been answered positively by the majority of interviewees. However, several constraints have been mentioned.

1) ..... possible, but not with the classical laboratory paradigms. Battery tasks should be made more complex. Measures should be gross in the sense that they are not restricted to measure an isolated process.
2) ......possible, but not with a limited number of tests that claim to cover the most relevant aspects of real life performance. It seems not possible to select a general battery that covers the large variety of human behaviour.
3) ......possible, but only after a detailed theoretical and empirical task analysis of the task under question. After specifying the major cognitive components it should be decided which lab tasks refer to real-life performance. Then the task remains to map the components to the theoretical model of task performance. This requires a process model dictating the elements involved, their interaction and possible ways to assess elements

and interaction. More should be known about the functional roles that skills and abilities play in the performance of real-life tasks. The process model should permit strategical freedom of the subject. It also should comprise the knowledge base and effects of practice.
3) ......possible, but only for sensory-motor tasks, not for complex tasks that involve command and control.
4) ......possible, but selection of subtest depends on the task under invesigation i.e. for the prediction of different tasks multiple batteries are needed.

A minority of interviewees were negative with regard to the aims of the project. They claimed that lab tasks are generally designed to study a special process in isolation and thus cannot have predictive value.

What other relevant methods were mentioned? Where are the current research needs?

a) performance measures

more status-oriented

- performance operating characteristics (POCs)
- measures of speed-accuracy trade off
- measures of decision bias (S/N ratio)
- rate measures (bits/second)
- more detailed analyses of errors
- measures derived from speed and accuracy (e.g. slope measures)
- dual-task performance (time-sharing)
- risk taking (e.g. measurement of safety margins)
- measures for the representation of knowledge

more process-oriented

- analysis of eye-movements
- analysis of verbal protocols ("thinking aloud")

b) subjective measures

- subjective estimates of workload
- state changes as indicated by subjetive measures
- similarity judgements

c) physiological indices

- state changes as indicated by physiological monitoring
- electrophysiological brain activity (e.g. evoked potentials)
- changes of pupil diameter

d) simulation methods

The interviewees agreed upon the fact, derived immediately from the above list, that the most obvious gap is in the assessment of control functions, i.e. the degree of systematic organizational planning of successively performed actions. A gap exists also with regard to tests that assess the integration of task components into task performance and the explanation of interindividual differences that might stem from different strategical preferences. Strategical aspects should be recognized as one of the major determinants of human performance and be assessed adequately by employing process models and process methodologies. The time has come to augment the standard repertoire by tasks that are designed to depict more the strategical aspects of behavior as they are relevant in performing real-life tasks.

Factor-analytic approaches may serve a good purpose in the exploratory or confirmatory phases of the research process. Due to their atheoretical nature they are useful for producing simple descriptions of the data. But the basic assumption of these models--that the human mind is a linear system--seems questionable. With regard to the aims of this project the modeling approach should be preferred.

2.2. TASK BATTERY REVIEWS

To assess the state-of-the-art in the area of standardized performance testing we have reviewed a number of widely used task batteries. (The selection does neither claim to be exhaustive nor to be representative in a strong sense.)

The following batteries have been included:
1. the BAT: Basic Attributes Test (US Air Force),
2. the CTS: Criterion Task Set (US Air Force),
3. the PAB: Performance Ability Test (US Army),
4. BBN: a battery developed by R.W. Pew et.al. for the US Air Force,
5. IPT: a set of information processing tasks developed by A.Rose,
6. TTP: the Ten-Task-Plan/ TASKOMAT developed by the TNO (Netherlands),
7. HAK: a battery developed by Häkkinen (Finland).

Fleishman's apparative setting and the results of the PETER project (Bittner, et. al. 1984) could not be included, because the authors did not send the detailed information we have been asking for before our deadline, July 1, 1986.

We have concentrated our review on the aspects of practical application and the reported theoretical background. Appendix B provides a detailed description of all tasks. A condensed overview is given in the table below.

GENERAL EVALUATION OF THE REVIEWED TASK BATTERIES

THEORETICAL BACKGROUNDS

The batteries reviewed here differ substantially with regard to their underlying theoretical frameworks. So far we have identified the following theoretical backgrounds:

CTS ---> MULTIPLE RESOURCE THEORY

BBN ---> GENERAL INFORMATION PROCESSING THEORY

BAT ---> FACTOR ANALYTIC APPROACH

TTP ---> ADDITIVE FACTOR APPROACH

For identifying the appropriate bases for a future battery, it seems necessary to review the theoretical frameworks found here and to evaluate which are the most promising with regard to the aims of this project. This should be one of the points for future work. Investigations should focus on the question whether the underlying framework is a broad enough basis for guaranteeing a reasonable prediction of performance in more complex real-life tasks. For instance, it has been repeatedly stated by major proponents of the additive factor logic that the method is only applicable in limited task domain (e.g., choice reaction tasks).

SELECTED SUBTESTS

Nevertheless the batteries do not differ that much in their choice of laboratory tasks. The following table where we have summarized the tests included in the batteries shows some surprising communalities :

T A S K	BAT	BBN	CTS	TTP	IPT	PAB
TRACKING						
one-hand		6	8	6		
two-hand	13					
TIME SHARING						
tracking + choice reaction	3					
tracking + memory		7	10	7		
tracking + dichotic listening		8				
DICHOTIC LISTENING		3		8		
SELECTIVE ATTENTION				3 4		1 2 10
VISUAL PROCESSING						
mental rotation	5	2	6			
embedded figures	10					
probability monitoring			1			
pattern recognition						7
PERCEPTUAL MOTOR SPEED	1 2 8			1		
MEMORY						
digit span		1				
Sternberg	6		3		4 5 8	
continuous memory	7		2	5		
digit recall						5
memory and visual search				2		
SEMANTIC PROCESSING						
Posner	4		4		1	
word meaning	11				2 3	9
Stroop		4				
sentence verification		5	7		6	4
Collins/Quillian						5
MATHEMATICAL PROCESSING			5			3 6
MOTOR PERFORMANCE			9			
RISK TAKING	9					
ACTIVE INTEREST INVENTORY	12					

(numbers stand for the running number of the test in the individual battery).

In all batteries reviewed here we find identical categories of tasks. The focus is on elementary perceptual-motor tasks, tasks testing elementary memory functions and semantic processing. The reasons for the striking resemblance between the batteries despite different theoretical frameworks should be more closely investigated. Either these tasks indeed cover the most relevant information processing functions or, in the other extreme, one battery has taken the other as a reference. Since none of these extremes appears to be true, the route from a theoretical framework to the choice of the actual task sample should be investigated.

Furthermore, the theoretical background and parameters of the individual task setting should be fully explored to get a deeper insight into the psychological processes involved in task performance. A next question concerns whether the tests are reliable and valid and whether they meet the necessary psychometric criteria. Do tests cover the most relevant aspects of human information processing in order to account for a major proportion of variance in the performance of a real life task?

Another striking resemblance relates to the fact that all the batteries reviewed here do not incorporate tasks that are supposed to tap higher mental functions like decision making or planning. In general, the more strategical aspects of behavior are neglected. Instead the focus is on elementary cognitive functions. It remains questionable whether a test device for performance in real-life tasks can afford to ignore the psychology of the 'mental executive' — a higher order process with the primary task of selecting and sequencing elementary cognitive functions.

In that context it should also be mentioned that all theoretical frameworks are related either to the classical psychometric approach with factor analysis as its principal methodological tool or to the information processing paradigm where it is taken for granted that every person does the test in the same way. There is increasing evidence that even in elementary paradigms the tasks are performed with different information-processing strategies. As long as these strategical aspects of behavior are not controlled and diagnostically evaluated a reasonable validity cannot be expected — especially not for the prediction of performance in real-life tasks. According to our view some extended effort should be spent on the design of new experimental paradigms and not on re-arranging already existing ones.

INTENDED PURPOSE OF BATTERY

It does not become clear in most batteries what are the basic intentions behind its construction. We may assume that in nearly all cases the assessment of reliable differences between persons has been the major aim. However, as Broadbent (see section on interviews) has pointed out the selection of subtests and their psychometrics may be radically different if one intends to measure the short-term effects of drugs or other stressors or of reliable personality characteristics.

3. METHODOLOGICAL ISSUES AND THEORETICAL OVERVIEW

3.1. METHODOLOGICAL ISSUES

ALTERNATIVE APPROACHES

The most direct approach to predicting on-the-job performance is a work-sample test. One simply allows the person to perform the relevant task using operational equipment, and evaluates that performance. Such a method is widely used in evaluating musicians and actors; in the entertainment field, it is called an "audition". Despite the appeal of this method, it is usually impractical for one of three reasons. First, there may be safety considerations that limit the use of operational equipment by persons of uncertain ability. For example, one would not wish to test the effects of drugs on pilot performance in a real airplane, even if laws and regulations permitted it. Second, one is often looking for aptitude—the ability to learn to do the job--rather than existing skill. Obviously we cannot obtain a work sample from an applicant who has not yet learned to do the job. Third, considerations of cost and equipment availability may preclude testing in an operational context. The high cost of operating real aircraft, tanks, ships, etc., make it impractical to conduct research on (e.g.) environmental stressors using operational equipment.

When the operational context cannot be used, for whatever reason, three principal alternative possibilities are evident; these are simulation, paper-and-pencil testing, and laboratory performance testing.

Simulation refers to the use of a functioning replica of the operational equipment/ situation for research, training, or selection (see also Section 2.1.). Simulators vary in fidelity from high fidelity, full mission simulators, in which the equipment and procedures are highly realistic, to low grade simulators in which only one or two operations of the real equipment are simulated. Simulations differ from standard laboratory tasks in that an attempt is made to faithfully recreate the function of the operational equipment. The principal limitation of a simulator, aside from costly initial developmentt, is that it is highly task specific, and must be redesigned for each change of application--often simply to perform the same task with new equipment.

Paper-and-pencil testing is usually aimed at testing a person's knowledge--either job knowledge or some more fundamental cognitive ability related to job performance . Knowledge testing is often quite effective in determining if a person has the proper job skills, even without asking the person to perform the job. Such tests tend to give a clear NO-GO for incompetent applicants. For example, a brick layer who doesn't know what a "bat" is, is clearly no bricklayer. There is the risk that a person may be able to "talk" a good job but unable to perform, and that is one limitation of this type of testing. However, where knowledge or cognitive ability is at stake, paper-and-pencil testing (or its oral equivalent) is the method of choice.

Laboratory performance testing has traditionally been used to test the effects of experimental variables on some relatively simplified performance such as simple reaction time, one- or two-dimensional tracking, pattern recognition, etc. Although it is often claimed that some variable will affect real-life behavior in the same way it affects a laboratory task, we have repeatedly stressed in this report that very little hard evidence is available to support this belief in the general case—and many of our interviewees have questioned whether generalization from the laboratory to real-tasks is ever justified. Despite this, recent results from some areas suggest that, with proper attention to detail, generalization from laboratory to the job can be supported (Sanders, 1984).

It is our position that these approaches are not redundant, and that each has its proper place in the field of human performance research. We see them as complementary, not at odds with each other. In particular, we see a standardized laboratory task battery as filling a niche that neither of the other approaches can fill efficiently. Compared to simulators, laboratory methods have the advantage of general purpose applicability and of being relatively inexpensive to implement and maintain. Compared to paper-and-pencil testing, laboratory methods provide a better ability to examine the perceptual-motor control and information-processing capabilities of the subject.

THE MEANING OF STANDARDIZATION

In the field of psychometrics, the expression "standardized" refers to a test for which the procedures for administration and scoring of the test are precisely defined. This means that the instructions, method of conducting the test session, test content, method of responding, and method of scoring are exactly the same for each individual being tested. Authorities differ on the question of norms, some saying that a test must have norms to be standardized, others saying that norms are not part of the definition. All agree, however, that norms are necessary if a standardized test is to be useful.

In the case of laboratory tasks, the notion of standardized testing means that the experimental procedure, task parameters, methods of responding, etc. must be precisely defined, so that the laboratory task is carried out in exactly the same way each time it is used. This has the merit that experiments conducted in different laboratories can be directly compared, and no allowances must be made for differences in procedure, stimulus materials, response manipulanda, etc., etc. Such standardization has obvious value to those who wish to use the results of research, if only because the number of contradictory research results will be reduced. The primary value, however, is that standardization makes possible the establishment of meaningful norms, against which the effects of new variables (e.g. drugs) can be assessed.

Standardization may cause some problems among individual researchers, who may resent being told that they must follow one specific procedure. It has also been argued that the regulation introduced by standardization may also act to stifle scientific creativity. These various merits and demerits must be weighed in deciding to promulgate any battery as the desired approach for research by any powerful funding agency. It is our opinion that the merits of standardization far outweigh other arguments.

In the course of this project, we have given some thought to the task elements that require standardization. In this section we wish to present a preliminary list of such elements for a limited selection of tasks which our reviews indicate as promising candidates, and on which meeting participants seemed to be in agreement. Neither the set of tasks nor the list of task elements is exhaustive; they should not be considered definitive, but only as representative of the decisions that must be made before finalizing any standardized performance testing battery.

LIST OF TASKS AND TASK ELEMENTS THAT REQUIRE STANDARDIZATION

1) TRACKING
a. type of display (pursuit, compensatory)
b. type of control (discrete, continuous, linear vs. rotary, number of dimensions)
c. type of input (step, ramp, sine, triangular, complex)
d. control dynamics (time lag, gain, control order)
e. preview
f. control-display compatibility (spatial, movement, conceptual)
g. spacing and predictability of successive inputs
h. single vs. multiaxis tracking
i. error feedback (accurate, inaccurate)
j. amount of practice

2) DUAL TASK PERFORMANCE
a. data limits (presence, absence, optimal loading)
b. structural interference (presence, absence, similar limbs, input organs)
c. resource-allocation instructions
d. modality specifity (same vs. different input systems)
e. response specifity (same vs. different response systems)
f. central processing specifity (verbal vs. spatial)
g. amount of practice

3) SPATIAL PROCESSING
a. paired vs. multiple comparisons
b. degree of rotation
c. angular disparity
d. axis of rotation
e. complexity of stimulus materials
f. familiarity of stimulus materials
g. kind of response ('same-different' judgement vs. telling from which perspective a standard stimulus is perceived)
h. testing 'spatial orientation' (e.g. cubes comparison) vs. testing 'spatial visualization' (e.g. paper folding tests, form boards) vs. testing 'spatial relations' (e.g. Cards, Flags & Figures)

4) CHOICE-REACTION PROCESSES
a. sensory modality (visual, auditory, tactual)
b. stimulus intensity/ contrast (low, high)
c. stimulus quality (intact, degraded)
d. stimulus content (verbal, signal lights, etc.)
e. stimulus similarity (similar, dissimilar)

f. set of alternatives
g. S-R compatibility
h. relative signal/ response frequency
i. time uncertainty
j. response execution
k. amount of practice

5) MEMORY SEARCH
a. target set size
b. target/ non-target category
c. single vs. repeated targets
d. consistent vs. varied target set
e. modality
f. type of target material (digits vs. letters)
g. amount of practice

6) LEXICAL AND SEMANTIC ENCODING (POSNER PARADIGM)
a. size of units (letters, words)
b. level of encoding (physical, name, category)
c. simultaneous/ successive matching
d. quality/ visibility of stimuli
e. meaningful vs. non-meaningful units
f. modality of presentation
g. interval between prime and stimulus

SELECTION AND EXPERIMENTAL USES OF A STANDARDIZED BATTERY

COMPARISON OF OF SELECTION AND EXPERIMENTAL APPROACHES

Research and applications in the field of personnel selection have universally been of the correlational type, and have focussed on the prediction of occupational success or on-the-job performance from predictor data. Thus the fundamental basis of selection is the examination of individual differences in predictor and criterion performance using correlational methods. Put simply, people who get higher scores on the predictor should get higher scores on the criterion if the predictor is valid. Differences in group means arising from differences in experimental variables across studies are generally ignored in correlational research as simply another kind of constant error that has no influence on existing correlations.

In contrast, experimental research is concerned with consistent differences caused by variation in experimental variables, and the focus is on group means. In experimental research, then, with rare exeptions, individual differences are simply a nuisance and are treated as experimental error. Only in recent years have correlational data been of interest to experimental psychologists, either in the growing acceptance of adjunct correlational techniques such as analysis of covariance or multivariate analysis of variance, or in the study of attribute-treatment interactions in human performance.

The predictable result of this history is that, for practical purposes, there are no normative data for laboratory tasks and few correlational data relating these tasks to either occupational success or real-life task performance. In addition, traditional reliabilities (test-retest, split-half, etc.) are rarely known for laboratory tasks. Only in the case of military aviation, where there has been a great deal of concern for predicting success in training, has there been much progress in relating laboratory tasks to real life performance, the classic case was the outstanding successs in selecting pilot trainees in the U.S. Army Air Forces during the second world war (Guilford, 1954). However, even in the extensive military research, many of the results are of limited generality or otherwise questionable.

The opposite face of the coin is that predictors for selection use have rarely been studied experimentally. Such predictors are often in the form of paper-and-pencil tests, usually testing occupational knowledge, general knowledge, or specific skills such as verbal ability, mathematical aptitude, problem solving, mechanical reasoning, or logical reasoning skills. While there would be little difficulty in researching the effects of experimental variables on paper-and-pencil test performance, it is our opinion that such research has not been fruitful, and would be of little use in developing a standardized task battery.

BATTERY VALIDATION FOR SELECTION

The process of validating a predictor test for selection purposes is straightforward, though not necessarily easy. One first chooses one or more criteria of successful job performance. The difficulty of finding "good" criteria is pervasive in selection work, and is known as the "criterion problem". Next, the task battery is administered. The scores on the various tasks are then correlated with each of the job criteria. For each criterion, optimum weights are selected for best predictors by multiple correlation technique. In the selection process itself, the applicants are ranked according to their composite score (based on the several best predictors), and the highest ranked applicants are selected for hiring, training, etc. There is no specific size of multiple correlation coefficient at which one says "the prediction is valid", although minima such as +.40 or +.50 are sometimes mentioned. Rather selection is a relative process; if your correlation coefficient is higher than that of other possible selection methods, the results are likely to be acceptable, even in the case of legal challenge.

As will be discussed in Section 3.2., Fleishman and associates have enjoyed substantial success in using the factor-analytic approach to develop sets of tasks for use in predicting on-the-job performance. Thus, although we are not fully committed to the selection use of a standardized battery, we have adopted a "wait-and-see" approach to this question; as Phase II develops, it should become clear whether a general purpose battery as envisioned by meeting participants and interveiwees will have any substantial utility for selection purposes.

USE OF A STANDARDIZED BATTERY IN EXPERIMENTAL RESEARCH

The validation of a standardized battery for research uses has a substantially different character from selection validation. The goal here is the demonstration that important variables affect performance on one or more battery tasks in the same way that they affect a real life task. For example, if drugs affect tracking performance in laboratory in the same way they affect manual control in an aircraft, we are justified in concluding that the tracking task is a valid predictor for manual control in the aircraft. Notice that individual differences are not the central issue; individual differences become important only in the case of attribute-treatment interactions, which are admittedly rare.

The crucial factor is that with a validated battery, one can then conduct research in the laboratory with reasonable assurance that the results can be generalized to a specific real life task. This not only permits substantial cost savings, but becomes critical when safety or other considerations prohibit needed experimentation in the operational environment. While a fully validated simulator (such as the ASPT facility operated by the Air Force Human Resource Laboratory at Williams AFB in Arizona, or the TNO simulator on manouvering ships at Soesterberg, NL) can also perform this function, there are few fully validated simulators, they are expensive to develop, operate, and maintain, and their operating time is essentially completely committed to other uses.

It is important to recognize that general purpose validation cannot be done in the early stages of development. The battery must be revalidated with respect to every real life task, including the case of the same task in different settings (e.g. freeway vs. city driving). Whether the future accumulation of results will permit broader generalization remains to be seen.

3.2. THEORETICAL OVERVIEW

THEORETICAL AND OTHER BASES FOR A STANDARDIZED BATTERY

RANDOM, INTUITIVE, AND PRACTICAL BASES

One possible approach is simply to generate a long list of candidate tasks, and to select randomly from these the desired number of tasks. Obviously the probability of selecting an effective battery by such a procedures is so small as to be negligible. However, a list of the problems with this approach may be instructive. First, one runs the risk of selecting several tasks of similar type, so that little is gained from using more than one of these. Second, crucial areas of tasks are likely to be omitted, so that the resulting battery covers only a limited portion of the relevant field. Third, because we know nothing about the composition of the tasks, the discovery that some important variable affects one or more of the tasks leaves us in the dark as to what aspect of the task is affected, and thus we will be unable to generalize the effect to other variables. Fourth, there are political problems to be faced in defending the battery from the onslaughts of those who are already committed to other task batteries, or from those who want to know the scientific rationale for inclusion of the various tasks.

Another approach is to choose tests on an intuitive basis, that is, choose tasks that one "feels" will provide a suitable spectrum of tasks to cover most needs. As one example, one could, with some justification say "we need a manual task, a pedal task, an arm-strength task, a leg-strength task, a visual task, an auditory task, etc., etc. To the extent that the battery constructor is gifted with an uncanny sense of intuition, this procedure might even work. However, most of us are not so gifted. Furthermore, it is clear that no two individuals will agree on the composition of the battery, if inclusion of tasks is to rest solely on individual opinion.

A third approach, which apparently has entered into the composition of more than one test battery, is practicality. If one already has the software for certain tasks, other tasks become less appealing; the same is true for tasks that are easy to administer, or for tasks with which one is thoroughly familiar and comfortable. There is, or course, necessarily an element of practicality in the construction of any battery by any approach. An eight-hour vigilance test or a task requiring a multimillion-dollar computer is not likely to appear in anyone's battery. But practicality should be assessed only near the end of the selection, rather than in the original screening of tasks.

Fortunately, there exist more rational approaches to selection of tasks for a battery. These are discussed in the following sections.

RATIONALE FOR TEST SELECTION

We take a strong position that one cannot simply throw together a set of tasks without looking more deeply to see what underlying skills and processes are being affected. It is tempting, of course, to include tests that have already been shown to be sensitive to the effects of some well known factor, for example, RT and alcohol. However, the existence of one or more known effects is no guarantee that the test will prove to be sensitive to other factors or classes of factors. Furthermore, there may be interactions that lessen or heighten the effects of some factor, for example, the well known synergistic effects of tranquilizers and alcohol, or the decrease of some effects with practice.

Thus, within any task, we need to examine in more detail the component processes or elements, so that it becomes possible to pinpoint the specific effects of an experimental variable. To do this requires a suitable model or theory of the task. Four major approaches may be identified, which provide useful models for establishing task batteries. These four are factor analysis, general information processing theory, multiple resource theory, and the processing stages model. In summary, the use of a suitable model or theory permits both the rational selection of tasks for inclusion and the detailed evaluation of the effects on these tasks of environmental, task, and internal bodily variables.

FACTOR ANALYSIS

A correlation coefficient is an index of the degree of relationship between two dependent variables, that is, two performance or response measures. The fundamental tenet of the factor analytic approach is that a significant non-zero correlation between two variables indicates the existence of a common underlying factor that, at least in part, determines the scores on both variables. Take, for example, the well known correlation between self-reports of cigarette smoking and the incidence of lung cancer. Obviously the self-reports of smoking do not "cause" cancer. Instead, the significant correlation indicates the existence of a common underlying factor, namely that the respondents did smoke cigarettes to the degree indicated, and that the smoking itself underlies both the self reports and the occurrence of cancer. Discovering and clarifying such common factors is the basic function of factor analysis

Factor analysis begins with a correlation matrix that shows the intercorrelations of many variables. It attempts to explain the patterns of intercorrelation by deriving a smaller number of factors that, in turn, would generate such a pattern. This provides a high degree of economy, because having to propose a separate common factor for every correlation between two measures generates unmanageable numbers with even small sets of variables. For example, with five tests, there are 10 possible intercorrelations, and in general, with x tests the number of pairs is (X*X-X)/2. In general, the number of factors is substantially less that the original number of variables; for example, if we have five tests of finger dexterity in our set of tests, we may very well find that a single factor accounts for the majority of the variance in all five tests.

The first product of a factor analysis is a factor matrix showing the "factor loading" of each test on each factor. This factor loading is an estimate of what the correlation would be between the test and a "pure" test of that factor, and indicates the degree to which the test contains that factor. Obviously factorially pure tests don't exist, but sometimes we find that one or another test is close to being a pure test of a factor. The next step is to identify each factor. If we find that a factor shows high loadings for tests of basic arithmetic, number series, mathematical reasoning, etc., we are tempted to identify this factor as "mathematical ability." With larger collections of tests, we may find that this factor really contains several component factors, such as number fluency, mathematical reasoning, and computational ability.

Factor analysts are divided on the philosophical basis of factor analysis. Some hold that there are "real" factors in nature, which are there to be discovered in the analysis. That is, there exist such abilities as mathematical, verbal, and mechanical ability, and that our analysis will uncover these factors if we are clever enough to include the proper tests in our battery. The "real factor" approach is necessarily based on some theoretical view of the structure or function of human mental processes and abilities. The opposite--and more popular--viewpoint is that a factor analysis is merely a way of looking at and summarizing data; that factors do not exist in themselves independently of our analysis, but are an interpretation of our data. However, the "data summary" people are willing to accept their own results as describing the state of nature, even though the "data summary" approach is largely atheoretical.

Prominent among the "real factor" theorists is J. P. Guilford, who has proposed a three-dimensional model called the "Structure-of-Intellect" (Guilford, 1977). The dimensions of the Guilford model are Contents, Operations, and Products. In this system, the Contents represent types of information that the organism can discriminate (visual, auditory, symbolic, semantic, and behavioral); Operations are the kinds of intellectual processing that can take place (evaluation, convergent production, divergent production, memory, and cognition); and finally, Products are intellectual outputs resulting from the organism's processing (units, classes, relations, transformation, and implications). This scheme was constructed <u>a priori</u> on the basis of Guilford's vast experience in the fields of intelligence, creativity, and performance measurement. To Guilford, the boxes in this three-dimensional model represent real entities, and the discovery of tests or test combinations to "fill" each of the boxes has been a major thrust of Guilford's research program. For example, the digit-span test (Wechsler, 1944) could represent an entry in the Auditory-Memory-Units box of the model.

In contrast to the real-factor approach, the data-summary approach starts with only a loose set of hypotheses about the nature of the factors to be uncovered. It is important to recognize that even the data-summary people do not start in a vacuum--without certain preconceptions, one would not know which tasks to study. One starts with observed consistencies in task performance, proposing abilities to account for these consistencies. Following this, the nature of the ability is further refined by careful factor-analytic correlational research. The goal is the selection of a set of tasks in such a way that each major underlying factor is represented in the task battery. This assures that experimental effects on each of the major performance factors can be evaluated.

A prominent figure in the development of task taxonomies on the basis of factor-analytic research is Fleishman (Fleishman, & Quaintance, 1984). Fleishman has carried out an extensive research project to identify major performance factors, to generate thereby a taxonomy of tasks, and finally, to create a set of rating scales so that the degree of each element of the taxonomy in a task can be reliably assessed. Fleishman used the result of factor analyses, both his own and those of others, to derive a list of 37 basic human abilities; these abilities ranged from verbal comprehension to control precision, and are tied directly to tests and laboratory tasks through factor analysis (Theologus, Ramashko, and Fleishman, 1973). Subsequently this list was expanded to 52 abilities, and published as the Manual for Ability Requirements Scales (MARS)(Fleishman, 1975; Schemmer, 1982). In the MARS manual, each scale is accompanied by a verbal explanation and behavioral anchors are provided in the scales themselves.

In summary, the factor analytic approach requires that one develop a set of tasks that cover the entire spectrum of relevant abilities. These tasks are related to the underlying ability factors by factor analysis, and scores on these factors can be derived from task scores by simple computations using the coefficients obtained in the analysis. Thus the effects of experimental variables on the basic underlying abilities can be determined. By inclusion of criterion scores from real-life tasks, one can also determine the relevance of experimental effects to real-life performance. Thus the use of factor analysis frees one from the intuitive quasi-inferences that derive from trying to interpret (e.g.) an experimental effect on reaction time in terms of real life performance such as automobile driving.

Both the real-factor and the data summary approaches have their advantages and disadvantages. First, a pure data-summary approach is not feasible, because, without some preconceptions, one would not know which tasks to include in the research and asnalysis. The data-summary approach is heavily subject to the choice of tasks, and it is conceivable that one may leave out an entire performance area unwittingly, by simply failing to include the relevant tasks in the research battery. In contrast, the real factor approach rests heavily on the investigator's wisdom in including all of the relevant dimensions and levels of dimensions in his original scheme.

There are, of course, more general concerns that apply to factor analyses of any type. The first is the indeterminacy of factor solutions; the final set of factors depends in part on the method of factor extraction, and perhaps more heavily on the adjustments that are made in the computations, called "rotations." This indeterminacy was so compelling to Guilford that he once (Guilford, 1954) suggested that, when all else fails, one should "rotate to psychological meaning" (p.509).

INFORMATION PROCESSING THEORY

Only a short remark on "information-processing theory": One frequently sees in the literature references to "information-processing-theory", as if there existed a single well defined theory qualifying for this title. In fact, such an expression refers to a large, amorphous mass of ideas and microtheories, many of which are mutually contradictory, and very few of which could genuinely be characterized as well developed theories. Included in this mass are such ideas as e.g. Broadbent's early "single channel" notion of input to the processing system.

For this report, we have chosen to single out two of the best developed theories for which there is now substantial empirical support, and which provide reasonable theoretical bases for the selection of tasks for a standardized battery. The two major approaches discussed below enjoy widespread support among experimental psychologists, although it is universally recognized that neither provides a complete account of all aspects of human information processing.

MULTIPLE RESOURCE THEORY

It has long been known that attentional capacity is limited. Even the ancient Greeks debated the question of whether or not it is possible to pay attention to more than one thing at once (Boring, 1950; James, 1890). With the growth in popularity of information-processing ideas, this became translated into the concept of limited information-processing capacity. The first popular idea was that the organism has a single, global, undifferentiated processing capacity, which is allocated--either through intermittent time sharing or through simultaneous apportioning--to the various tasks demanding attention. When all resources are being utilized, increases in the resources (capacity) devoted to one task were necessarily taken away from other tasks, causing a decrement in performance of the other tasks (Moray, 1967; Broadbent, 1958).

The notion of a single pool of attentional or information-processing resources has proven impossible to sustain. Wickens (1984) has pointed out several reason for this. First, in numerous cases, the interference between two tasks is related, not to their difficulties, but to their structures. For example, keeping pressure on a stick interferes much more with tracking than auditory signal detection, even though the detection task was judged much more difficult. Presumably this is cause by the greater structural similarity between between maintaining stick pressure and tracking.

Second, certain combinations of tasks demonstrate "difficulty insensitivity." This means that increasing the difficulty of one task, which should consume more capacity or resources, does not affect performance of a second task. An example cited by Wickens was a case in which three different levels of complexity of a discrete numerical response task interfered equally with performance on a tracking task.

Third, two tasks that obviously demand substantial attention can sometimes be time shared perfectly. Wickens describes cases in which skilled pianists can time share sight-reading music with verbal shadowing without decrement to either task; the same result was found with with skilled typists transcribing written messages.

Although some of these results might possibly have alternative explanations, such as automatizing of one or another task, various investigators (Navon & Gopher, 1979; Sanders, 1979; Wickens, 1979) have proposed a "Multiple Resource Theory" (MRT). This theory proposes that instead of a single undifferentiated pool of processing capacity, there are several independent sets, or "pools" of resources. The fundamental tenet of the theory is that if two tasks draw heavily on the same resource pool, they will interfere with each other; if they draw on separate resource pools, there will be no mutual interference when the tasks are performed together, and changes in the difficulty of one task will have no effect on performance of the other. In sum, two tasks will interfere with each other to the extent that they share the same resource pools. It is important to note that a given task draws from its resource pools regardless of whether or not it is being performed in conjunction with other tasks; the dual task methodology is used only to determine which pools are shared and which are independent. Thus MRT is not inherently tied to dual task methodology.

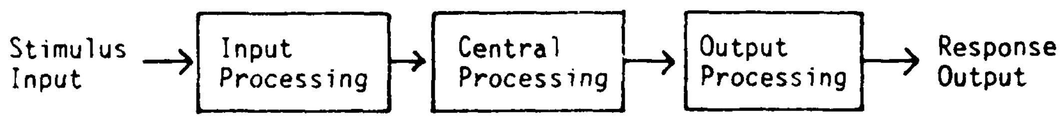
MRT, as proposed by Wickens, conceptualizes resource pools as lying along three dimensions, Stages (encoding, central processing, and responding), Modalities (visual and auditory), and Codes (spatial and verbal). There is also a suggestion of a fourth dimension of manual vs. vocal responding, but this is not specified as being separate from the spatial vs. verbal dimension. One interesting point of the model is that Wickens conceptualizes both encoding and central processing as representing the same resource pool. This is a major difference from the "stages" formulation, to be discussed below. in which input processing and central processing are considered to be independent processes. Although Wickens is aware that new data might force the inclusion of more levels of any of these dimensions (such as needing to include a tactile or kinaesthetic modality) he warns strongly against allowing the model to expand indefinitely.

In using MRT as the basis of constructing a test battery, the goal is to select a set of tasks so that each resource pool is represented in the battery. Furthermore, when a dual task test is included, one must pay strict attention to whether the two tasks share the same or separate resource pools. The obvious difficulty in this approach lies in the limitations of the model. As mentioned above, it is not difficult to propose resource pools not included in the model--tactile or kinaesthetic modalities, for example. It is conceivable that there exist environmental or drug effects on these modalities that are only minor for visual and auditory modalities, and might thereby pass unnoticed. The same considerations arise when interpreting results obtained with such a battery. To the extent that the set of tasks covers all relevant resource pools, an effect of some stressor should appear at some point in the battery. With a sufficient number of tasks of varied kinds, it should be possible to pinpoint which pool or pools is being affected. This in turn should allow a fine-grained analysis of the effects of the variables in question.

The MRT is still somewhat controversial among psychologists. Some, for example, have taken the position that there exists an undifferentiated central resource pool accompanied by numerous independent dedicated resource pools (Neisser, 1976). Others feel that there is too much danger of proliferation of resource pools, so that, in the end, one is left with an unwieldy concept of little or no practical utility. Finally, others have questioned the notion of complete independence of pools, with the implied notion of separate processing systems for different kinds of information. It is, in fact, difficult to conceptualize a system in which visual attention to spatial and verbal information operate simultaneously and independently of each other. However, the data seem to support the existence of a multiple resource system much like that proposed by Wickens. - for a more detailed discussion of this issue, and a review of the relevant literature. see Appendix 5.4.2.

STAGE THEORY

As the information-processing approach to complex human behavior grew in popularity during the period from 1950 to 1970, psychologists began to discard the notion that complex behavior could be broken up into separate islands labeled "perception," "learning,""motivation," etc. In its place, there arose a conceptualization that behavior could be best understood by examining the flow of information through the organism, from sensory input, through detection, encoding, and recognition, to central processing, response selection, and response generation. Simple linear models were proposed for this process, most of them like the following:



During the late 1960s, it was proposed that the processing of stimulus input takes place in a series of non-overlapping independent stages (Sternberg, 1969). Sternberg's work was concerned primarily with reaction time, in particular, the time required to search memory and decide if a stimulus probe is a member of a pre-memorized set of materials. Sternberg proposed that reaction time is simply the sum of the times required by the individual processing stages; the notion of independence implies that two variables that influence a common processing stage will interact. Correspondingly, if two variables affect only different stages, their effects will be additive, that is, they will not interact. This is the source of the title "additive factors method" often associated with this approach to the study of information processing.

The initial version of processing stages theory included only three stages--perceptual encoding, response selection, and response execution. The central portion, response selection, included such processes as memory search and decision to explain processing in a traditional choice reaction. In more recent years as many as six stages have been proposed (Sanders, 1980). As with Multiple Resource Theory, authors have warned against allowing the number of stages to proliferate.

The use of the processing stages model for selection of a task battery poses a difficult problem. As Sanders (1984) has pointed out, the processing stages model has shifted the emphasis from comparing tasks to comparing variables. The scope of the processing stages approach is, in fact, limited to the variety of multiple choice reaction and memory search tasks. This implies that one or two tasks will suffice for the assessment of processing-stage effects in a standardizxed battery. The actual effects then must be assessed by studying the effects of known and standardized variables on performance. For example, it is well known that variations in stimulus quality affect aspects of stimulus encoding; "fuzzy" stimuli require longer to encode, and raise overall reaction time. One can imagine that a drug that mimics this effect, by reducing the clarity of vision, may also raise overall reaction time, without simultaneously influencing the effect of number of stimulus alternatives (which affects the central processing stage)

Sternberg (1969) and others have pointed out that there is a distinction between processes and processing stages. Processes take place within stages. For example, memory search and the yes-no decision concerning the stimulus both take place in the central processing ("response selection") stage. This suggests a further elaboration of battery development. Perhaps one should select not only tasks and variables that will distinguish between stages, but tasks that can be clearly tied to one or another process within stages. It is not entirely clear how this can be done rigorously, avoiding the need for intuitive judgement concerning the existence or non-existence of processes.

In summary, a processing stages model cannot provide a complete basis for a task battery, because it deals with a limited set of tasks which clearly do not cover the entire spectrum of human performance. Instead, for those reaction tasks that are included, the processing stages model provides a means for a finer grained analysis by isolating the effects of important variables to a specific stage of processing. Like the other theories presented here, the processing stages model and the additive factors method associated with it are controversial. However, as a practical matter, stage analysis holds great promise for helping establish at least one major portion of a standardized task battery. More detailed discussion of stage analysis and the additive factors method, together with a review of the empirical literature on choice reaction time and memory search are found in Appendices 5.4.4. and 5.4.5.

REFERENCES

Boring, E.G. (1950). A history of experimental psychology (2nd ed.). New York: Appleton-Century-Crofts.

Broadbent, D. (1958). Perception and communication. Oxford: Pergamon Press.

Fleishman, E.A. (1975). Toward a taxonomy of human performance. American Psychologist, 30, 1127-1149.

Fleishman, E.A., & Quaintance, M.K. (1984). Taxonomies of human performance. New York: Academic Press.

Gopher, D., & Sanders, A.F. (1984). S-Oh-R: Oh stages! Oh resources! In: W. Prinz, & A.F. Sanders (Eds.), Cognition and motor processes. Berlin: Springer.

Guilford, J.P. (1954). Psychometric methods (2nd ed.). New York: McGraw-Hill.

James, W. (1890). Principles of psychology. New York: Holt.

Moray, N. (1967). Where is attention limited. A survey and a model. Acta Psychologica, 27, 84-92.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. Psychological Review, 86, 214-255.

Neisser, U. (1976). Cognition and reality. San Francisco: Freeman.

Sanders, A.F. (1979). Some remarks on mental load. In: N. Moray (Ed.), Mental workload: Its theory and measurement. New York: Plenum Press.

Sanders, A.F. (1980). Stage analysis of reaction processes. In: G.E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Sanders, A.F. (1984). Drugs, driving, and the measurement of performance. Paper presented to the First International Symposium on Drugs and Driving. Vinkeveen, The Netherlands.

Schemmer, F.M. (1982). Development of rating scales for selected visual, auditory, and speech abilities. (ARRO Final Report 3064). Washington, DC: Advanced Research Resources Organization.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. Acta Psychologica, 30, 276-315.

Theologus, G.C., Romashko, T., & Fleishman, E.A. (1973). Development of a taxonomy of human performance: A feasibility study of ability dimensions for classifying human tasks. JSAS Catalog of Selected Documents in Psychology, 3, 25-26, (MS. No. 321).

Wechsler, D. (1944). The measurement of adult intelligence. Baltimore: Williams & Wilkins.

Wickens, C.D. (1979). The effects of divided attention on information processing in tracking. Journal of Experimental Psychology, 1, 1-13.

Wickens, C.D. (1984). Processing resources in attention. In: R. Parasuraman & D.R. Davies (Eds.), Varieties of attention. New York: Academic Press.

4. CONCLUSIONS AND RECOMMENDATIONS

1. It is encouraging that most interviewees were positive toward the idea and the feasibility of a standardized task battery for human performance research, despite the fairly general consensus that simple laboratory tasks cannot be supposed to be highly predictive with regard to complex real life tasks. The lack of predictive power seems to apply irrespective of the envisaged purpose of the battery--whether personnel selection, assessing environmental effects on performance, or predicting system performance. The notion of standardization has considerable appeal, if only to permit comparisons between the results of different laboratories and a more systematic attack on the question of the ultimate predictive value of various laboratory paradigms.

2. It is interesting that there are considerable commonalities between tasks in the various existing batteries (see table in section 2.2.), and among the preferences expressed in the interviews. The following is a minimal list of tasks on which there was widespread agreement:

a) tracking--preferably a critical instability task (see review 5.4.1.),

b) dual task performance--in which one of the tasks is usually tracking, combined with a discrete task that often has a short-term retention component (see review 5.4.2.),

c) visual processing--with an emphasis on mental rotation, pattern recognition, or embedded figures (see review 5.4.3.),

d) choice reaction processess--preferably tasks in which the effects of some critical variables are determined (see review 5.4.4.),

e) short-term retention measures--Sternberg's memory search and continuing memory tests are the most popular (see review 5.4.5.),

f) linguistic and semantic processing--mainly relating to Posner's matching paradigm and Baddeley and Hitch's sentence verification task (see review 5.4.6.).

It should be noted that these paradigms also appear in a recently proposed list for a tri-service battery of performance tests. Again, they show a considerable convergence with the list that appeared in the proceedings of the Paris meeting on standardization (27-29 May, 1986).

3. It should be clear that a decision to limit the battery--at least for the immediate future--to these types of tasks is merely a first step. The next issue concerns the determination of the optimal set of parameter values and other characteristics of each individual task. On the basis of the reviews of the literature, we have prepared a list of some major parameters needed to be set (section 3.1.). Choosing a collection of tasks is relatively straightforward; determinig their final shape requires a considerable amount of additional thought and consideration, and experimental tryout as well.

4. It is widely felt that a task battery consisting of tasks as outlined in Conclusion 2 puts too much emphasis on perceptual-motor performance, and fails to give enough emphasis to strategical elements of performance. It should be added that additional items proposed for the tri-service battery include vigilance, pattern comparison, code subsitution, time estimation, interval production, and Stroop interference, which also

tap classical perceptual-motor performance. A more careful inspection of items in the various task batteries shows the same trend with the possible exception of "risk taking" in the BAT and "probability monitoring" in the CTS. In the probability-monitoring task, subjects decide about deviations from randomness, which presumably does not carry beyond perceptual processing. For the risk-taking test, subjects maximize gains by opening a number of boxes. Each box delivers a certain financial benefit, except for one box which inflicts a considerable loss. Although it cannot be denied that the risk-taking contains cognitive and strategical elements, it is still closely bound to classical decision research.

It is probably not surprising that cognitive and strategical aspects are not emphasized in existing batteries, since well researched experimental paradimgs underlying such tests are not yet available. Consideration and further development of such tasks is a major issue for future research.

5. Conclusion 4 can be extended by noting that, as yet, our project does not propose a major orgnisational scheme or taxonomy. Yet it seems impossible to do a proper job of developing a battery unless such a scheme is available to insure that nothing is left out.

The principal effort toward formulating a task taxonomy is undoubtedly contained in the research of Fleishman and his coworkers (Fleishman & Quaintance, 1984). Although the majority of the interviewees rejects the correlational approaches advocated by Fleishman in favor of the information-theoretic approach--as also exemplified in the present report by the way of the theoretical reviews underlying the tests of the various batteries--there is an obvious need for better communication between both approaches. This concerns the communality between the respective task batteries as well as the task analysis approach for characterization of real life tasks.

6. This report has relatively little to say about the crucial question of how to relate the laboratory-based tests of any task battery to on-the-job performance. Many interviewees felt that laboratory tasks have little predictive validity because of (a) their context reduced nature, (b) the failure to account for interactive effects as observed in more cognitive skills, (c) the failure to include strategical effects, and finally, (d) the low level of practice that is accomplished in laboratory tasks (see section 2.2.). In fact, the much discussed predictive validity of laboratory tests is more a hypothesis than an established truth. However, if this feeling were generally valid, and if a task battery would ultimately fail to reflect important aspects of real life performance, the trade of investigating human behavior in the laboratory would make little sense (see Sanders, 1984, for more extended reflections on this issue). Yet the situation does not appear to be that dim, and we do not share the pessimism evident in some of the interviews, (see also Broadbent's remarks in section 5.2. for a more positive outlook).

Clearly a major task for future research, and for this project specifically, consists of validating standard laboratory tasks, using real life settings that permit the development of reasonable criteria. One major difficulty in that effort will obviously be that real life performance validation suffers from the usual criterion problem, which renders the process quite difficult. However recent developments have been encouraging;

car driving in experimental cars that allow measurement of various behavioral and psychological parameters is a case in point. This recommendation is fully in line with the back-to-back experimentation procedures suggested by Gopher and Sanders, 1984.

7. In the near future it will be time to decide whether we wish to pursue a fixed-content or a "laundry-list" (variable-content) approach to battery development. A good case can be made for either. A fixed-content battery is one in which a specific set of tasks is defined, and which purports to cover the entire field of interests. The classical examples of fixed-content test collections are the Stanford-Binet and Wechsler intelligence tests. The variable-content approach, in contrast, is one in which a much larger set of tests is standardized, and one selects a subset of tests for each application. Such an approach provides a greater flexibility in that tests can be added as desired, and that one needs not administer tests that have been shown not to predict in a give situation. Many of our interviewees expressed a concern that a single fixed-content battery would never be able to cover a sufficiently broad spectrum of human performance to be of general use.

We would be most comfortable with something like an "enzyclopaedia" or handbook of standardized tasks, complete with norms for each task. Subsets of these tasks could then be validated for specific real life tasks. Such a handbook could be expanded as opportunities arose, for example new tasks could be added as they are invented. Clearly any fixed-content battery put together 20 years ago would suffer from omission of many procedures that are considered useful today (e.g., Shephard's "mental rotation" or Sternberg's memory search).

We envision this proposed handbook as something similiar to the U.S. Pharmacopoeia, only with a list of "recognized" standard procedures rather than drugs. The only problem we foresee is the resistance from experimental psychologists, who are likely to feel that someone is dictating to them how to conduct their research. Despite this, we have felt for some time that something like a handbook of performance measurement is badly needed, if only to reduce the frequency with which our collegues have to "reinvent the wheel"; while on sabbatical at NASA, one of us (Haygood) was struck by the fact that every new project brought forth the cry that "we have to deal with the measurement problem".

We have been impressed by the fact that several exellent task batteries already exist, and that identifiable deficiencies in these batteries would be easy to rectify by adding more tests. It is clear that the value of the current project lies, not in developing yet another battery (just like all the rest, albeit a bit more complete), but in examing the fundamental premises on which most batteries are contructed. Specifically, we recommend exploring in depth the variable-content or "Pharmacopoeia" approach during Phase II. This recommendation would clearly lead to an expanded candidate list of tasks, and would viviate the destructive effects of discovering that one or another task was inadaquate for the purposes of a standardized battery.

REFERENCES

Fleishman, E.A., & Quaintance, M.K. (1984). Taxonomies of human performance. New York: Academic Press.

Gopher, D., & Sanders, A.F. (1984). S-Oh-R: Oh stages! Oh resources! In: W. Prinz, & A.F. Sanders (Eds.), Cognition and motor processes. Berlin: Springer.

Sanders, A.F. (1984). Drugs, driving, and the measurement of performance. Paper presented to the First International Symposium on Drugs and Driving . Vinkeveen, The Netherlands.

5. APPENDICES

5.1. PROJECT EXPLANATION FOR INTERVIEWEES

Prior to the interviews the interviewees were briefed about the project by reading the following explanation:

STANDARDIZATION OF PERFORMANCE TESTS: A Research and Development Project

A major problem today is that researchers in human performance have used a variety of experimental methods and tasks. Even when the task is ostensibly the same (e.g., the Sternberg memory-search task), experimenters have used different task parameters, equipment, stimuli, instructions, etc. This lack of standardization has created several problems for those who wish to use results for practical decision making. For example, there are no norms for the various experimental tasks. In addition, there is a widespread complaint that the methods and tasks used in the laboratory are so simple and artificial that they have little or no applicability to real world tasks. We are starting a project (for the U.S. Air Force) dealing with the possibility of developing a standardized battery of performance tests.

Our immediate task is to review the literature and talk to active researchers and theoreticians in the field of skilled human performance. For this reason we are interviewing a number of people who work in this field. The information gained here will help guide our further efforts.

The ultimate purpose of this project is to establish a collection of standardized laboratory methods for studying and measuring human performance, and to clarify the relationships between these methods and the components of important real world tasks.

Ideally, results should make it possible to provide "standard" versions of many tasks, so that

(a) Experiments conducted in different laboratories, by different people, at different times, and with different subject populations can be compared and integrated.

(b) Norms can be established for each task, including not only norms for various task parameters, but also norms for different types of subject (age norms, sex norms etc.).

(c) Assuming that it is possible to perform meaningful component analysis for real-life tasks, the relationships between such components and laboratory-task performance, if any, can be established.

(d) The theoretical basis of skilled human performance can be further developed, including questions of linear-additive models, multiple vs. single resource pool models, parallel vs. serial processing, etc.

Our first concern is that of feasibility. Can we, in fact, develop such a standardized battery? And if so, how can we best use the work of others as a foundation for this project? What relevant sources have we missed?

We are, of course, also concerned with the desirability of such a standardization. Some people feel that general acceptance of specific methods of doing certain kinds of research would have a constraining/limiting effect on research creativity. However, it is not our intent to develop anything like a "skilled performance I.Q. test", and we are certain that the existence of this battery will not constrain the creative development of new laboratory methods.

The success of the project should provide some practically useful benefits. In particular, it would be useful to have a standard set of tasks to assess the effects of stressors such as drugs, noise, and lack of sleep. In addition, it would be valuable to have a standard method of testing the perceptual-motor load of many real-life tasks. Finally, an understanding of the relationship between real-life task components and the standardized battery of performance tests should be useful in personnel selection.

5.2. INTERVIEWS

LIST OF INTERVIEWEES

Dr. Donald Broadbent
Written Comments on discussion topics for 'Standarisation of Performance Tests'

Preliminary

I ought to explain the difficulty I had regarding the whole list of topics; namely that it was not clear what purpose was envisaged for a battery of performance tests. There are three broad classes of purpose, and the answers would be quite different for each of them. First, the battery of tests may be used to assess differences between individual people. Second, they may be used to assess the effect of some environmental conditions such as drugs, anoxia, sleeplessness, and ciradian rhythm. Third, they may be used to assess the impact of some proposed new task or sub-task on a total complex of performance; for example, whether the use of verbal annunciator systems for communicating information to the pilot will help or hinder other activities in the cockpit. The requirements of a task battery for these three needs would be different, for the following reasons. To assess individuals, one wishes to find measurements that are extremely stable for that same individual; test-retest correlations should be high, inter-individual variation large, and of course validity in this sense means a high correlation between the individual differences in the test and individual differences in the criterion performance. To assess environmental changes, however, just the opposite is true. We want tests which fluctuate markedly when the environment changes. Ideally furthermore we want differences between individuals to be small, so that the theoretically preferable separate groups designs can be used for comparing environmental conditions. Both these factors mean that test-retest correlations will be extremely low, and prediction of individual differences in a criterion task from test performance will also be low. Validity in this case means that the group average should change in the same direction in the real task as it does in the test, which is a quite different requirement from the first purpose. For the third purpose, one would like tests in which the individual and environmental components of variance are low, in order to increase the power of experiments. On the other hand, the impact of changes in job design depends on the exact functional mix of tasks being used, since for example the speech annunciator may have quite different impact on other speech tasks and on visual tracking tasks. It then becomes very important that the tests contain a representative sample of tasks in each of the processing domains; which is not necessarily true of tests used for the other purposes. Because of this difference of requirements, I should have thought that the answer to all the discussion topics would be different depending on one's interests; and that the realistic aim would be for three batteries of tests rather than one! I have not recapitulated all this under each heading.

1. This group uses a very wide range of methods. They fall into four main classes.

(a) Methods for assessing effects of environments. These are intended to test relatively isolated functions, and to cover a range of such functions because it is often unclear which ones are liable to be impaired by some particular environment. There is a trade-off between the time taken by the tests and the number of different functions that can be tested; the most common group of tests is a serial reaction time, syntactic reasoning, sentence verification from common knowledge, and vigilance/running memory (prolonged concentration). All these are microcomputer based and reasonably portable. Other tests less used as yet include the Eriksen technique for measuring effects of distractors (as discussed in may Aachen paper), spatial non-verbal memory test, and tests of assignment of words to categories.

(b) **Methods of analyzing particular detailed function with traditional laboratory designs.** Most of the interest here lately has been in attention; in addition to the Eriksen techniques, mentioned above, we use the monitoring of rapid serial visual presentation lists in search of a target, and lexical decision. In the field of memory, we tend to use serial presentation of short lists under varying conditions of recall order, type of stimulus, and activities intervening between presentation and recall. POC analysis of the results is extremely illuminating.

(c) **Unconventional laboratory tasks of a simulation type: aimed at control processes.** These include computer simulated interactions with other persons, simulations of running an economy, or managing a factory; in some cases, playing fairly complex video games. Most activity in this area has concentrated on the relation between explicit reportable knowledge and successful performance that cannot be described verbally.

(d) **Questionnaire studies.** These are either characterisation of jobs on various established scales (such as those of Karasek), or self-report measures of current state, or chronic characteristics such as liability to cognitive failure. Currently these questionnaires make the main bridge between laboratory and the field, as they are used in both situations.

2. Particularly useful methods.
This question is very difficult to answer; for my own purpose, naturally I regard the methods we are using as the best both theoretically and in terms of generalised ability. They might well be unsuitable for people with rather different interests.

3. Alternative metrics.
In one sense, clearly no information can be obtained about human beings except in terms of what they do and the time at which they do it. However, several measurements may be combined in useful ways; the value of performance operating characteristics has already been mentioned, and speed accuracy trade-off functions are also important although we ourselves have not used them much. Similarly, we keep an eye open for relatively low frequency rhythms of performance, which might indicate the effect of a higher order monitoring control; but have not yet found them working.

4. I am afraid I do not accept the suggestion that tests have a low validity for real-life. My own experience is mostly in the second of the three areas; in that area, I know a number of cases in which validity has been assessed in real-life, and has always been found to be satisfactory. I

know of no case where an assessment has been made and found unsatisfactory. This is despite the fact that armchair arguments of abstractness et cetera were used against laboratory demonstrations of the impact of alcohol on car driving, or radar performance; more recently, of marijuana and valium. So far as pilot performance goes, Nicholson's reports of the use of benzodiazepines to control circadian rhythm problems in operational aircrew seem to me very sound validation of the laboratory test that gave rise to the methods used. With regard to the third of the three aims, there are classic validations from accident records of laboratory tests either of lever positioning (Fitts in the Berlin airlift) or of instrument displays (the three-point altimeter). The one area where I might admit some lack of validity is in the first aim, selection of able individuals; it is well admitted that the prediction of pilot performance from existing selection batteries is bad. One possible reason for this is the low variance of ability amongst people admitted to flying training. If this is the explanation, it is insuperable. It may also be however that there is an extra function that needs assessment.

5. The gap in the existing set of tests that are available is any measure of control functioning. That is, I do not know a satisfactory measure of the reliability with which somebody will move from one sub-task to another in a complex environment including a number of such tests. Contemporary computer techniques make it possible in principle to produce such a test, which was not so relatively recently; it is perhaps the area of development which should be most encouraged. In the realm of individual differences, this shows up in the rather simplified form of the debate over "time-sharing ability". It is more than that, as the tests normally used for time-sharing are very simple ones performed during the same broad periods of time. I am thinking much more of the degree of systematic organisational planning of actions performed successively, which is certainly needed in many real-life situations.

6. The breaking down into components of real-life tasks.
My spontaneous answer here is "this is possible to a high degree". However, there is no very clear scale for measuring the degree; doubtless it could be improved. I would however argue that in general it is possible to assess a task for the extent to which visual or auditory perception is involved, detailed manual control or speech, maintenance of alertness, use of working memory, and so on. As noted previously the main weaknesses is any test of the component (which is logicaly necessary) that keeps the various subsidiary components in balance.

7. Feasibility of a standardised battery.
See the introduction; it would need to be a different battery depending on the purpose for which it was used. It would also need to be a much larger battery than would normally be employed for any particular application.
(a) Inclusion of tests; at the very least, all those I have mentioned should be in, and probably a number of others.
(b) Factor analytic approaches are extremely useful for producing simple descriptions of the data relative to hypotheses that have already been formulated. Unless however a test of a particular function has been included in the battery, factor analysis will naturally not show it up. I also object in principle to factor analysis, as opposed to other

mathematical techniques, because even for functions that have been covered the exact factor solution will change depending on the other tests in the battery. It should be remembered that Fleishman's approach is directed primarily towards the individual difference problem; I would not accept that analysis of the correlation across individuals necessarily sheds any light whatever on the development of tests for environmental conditions, nor for the evaluation of changes in the sub-tasks. This refers back to the introduction again.

8. I would certainly think that a broad enough battery of tests can be devised; there are certain areas of weakness, such as the testing of control processes already mentioned.

9. I am not quite sure what is meant by 'skill categories' in this question; my natural inclination would be to have tests that measure resources of the individual, that is the quality of certain lasting representations, and the efficiency with which processes transform one representation into another. It is also necessary of course to classify the tasks, as in the classic distinction of open and closed skills. A skill that requires continued feedback from the environment makes use of different resources from one that can be executed in a ballistic way once the conditions for the action have been observed. It may frequently be that there are certain skills that place no load on working memory, and so on. Hence, it is necessary to distinguish categories of task in terms of the requirements demanded of the person, and categories of resource in terms of what the person can contribute to these tasks. My problem is that I am not quite sure which of these points the question was emphasising.

Dr. Susan Chipman
Office of Naval Research, Washington D.C.

The interview with Dr. Chipman was not based on the list of questions given in the introduction. It was the primary purpose to explore what other agencies are engaged in a project like this. However, in the course of the discussion, the following points were made with regard to the original list of questions:

Dr. Chipman points out that measures of working-memory capacity as developed by Meredith Daneman et. al. could be of some value in the assessment of human performance (e.g. Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. JVLVB, 19, 450-466)

It would be worthwhile to investigate the role of metacognition in the process of task performance (cf. Sternberg's metacomponents). There seems to be a steadily increasing interest in the psychology of interindividual differences - especially with regard to different strategies to perform a task.

She is a bit sceptical with regard to the possibility to break down a complex task into distinct components. Tests would only pick up a tiny fraction of what really happening. This would explain the low validity that is usually observed. Tests do not pick up the control or coordination exerted by a mental executive that is directing the component processes in the performing a complex task. Any real-life task is supposed to be complex. In that sense the contribution of factor-analytic approaches cannot be very substantial since these approaches do not take into account a mental executive. Thus they can never present a complete picture of the human mind.

In general there is a fair probability that breaking down a task into its components may be achieved, but if no detailed theoretical knowledge about the task exists there is no guarantee of a succesful approach. Obviously there is little agreement among researchers what the "real" components are.

Dr. Allan Collins
Bolt, Beranek & Newman Inc., Cambridge MA

Dr. Collins is working in the areas of semantic processing, use of computers, and education. His research interests are both applied and basic. At the moment he is involved in research on mental models in physical systems and the design of computerized teaching systems.

In his research he has employed almost every kind of experimental method. More recently he has focused on protocol methods and on discourse analysis.

To him errors are not a metric per se. Errors can have very different causes which should not be intermingled. Therefore a more thorough and qualitative analysis of errors could yield some better insights into cognitive malfunctioning. He mentions "repair theory" (Cognitive Science 1980-81) as a prominent example here.

He sees some problems with regard to the extrapolating from test scores to real-life performance because of the lack of face-validity. Most experimental paradigms are not aimed at evaluating all the variables that affect performance. The main reason here is that tests of isolated abilities or skills never incorporate the interaction effects when these skills have to be combined in a task.

With regard to question 4 he argues that laboratory tasks can never have high validity with respect to the prediction of real-life performance simply because they "cannot do the job". Real-life task are by far more complex and involve interaction effects between various elementary cognitive processes. Laboratory task are designed for the purpose to study a phenomenon in isolation. Therefore real-life tasks and laboratory task represent endpoints of a continuum. A low validity therefore must be expected.

The generalizability of laboratory tasks could be improved by drastic changes of these tasks. They should be made more complex and it should be clear which cognitive processes are involved and how they interact.

This means that the reliable assessment of components of a real life task is the critical thing that has to be achieved. Dr. Collins mentions some research activities of Earl Hunt and Robert Sternberg that point in the same direction.

Dr. John Frederiksen
Bolt, Beranek & Newman, Inc.

Dr. Frederiksen is working in the area of cognitive psychology, especially in reading research and in the componential analysis of skills. His present basic research interests are covariate modeling, decomposing skills of reading, skills interaction in reading, and instruction and training. He has also worked on the teaching of complex skills and intelligent tutoring systems.

In his reading research Dr. Frederiksen has mostly applied methods that are specific for reading research (pronounciation tasks, lexical decision tasks, reading span, reaction time tasks). These task were all theoretically motivated. Dr. Frederiksen emphasizes that in his domain standard laboratory tasks would not do the job because they might not be related to specific aspects of reading.

The standard repertoire of experimental metrics should be augmented by tasks that depict more strategical aspects of behavior. Here he mentions the scores derived from video-game play (knob-usage).

For him the reasons for the low validity of performance tests lie mainly in the integration and coordination of skills. He mentions the problem caused by the "automaticity" of tasks with increased practice that might change the factorial structure of underlying skills drastically and problems caused by the interference of skills in certain tasks.

A better validity can only be achieved if more is known about the functional roles that a skill has in the whole task. Also the influence of strategic differences should not be underestimated. In that context he mentions the reports of Andy Rose for the Office of Naval Research as an example. The best procedure would be to carefully study the task, get an idea what cognitive components are involved in performing the task, develop experimental paradigms for assessing these components. A strong emphasis is put on so-called "top-down-analyses" of human task performance. The set of predictors, however, should also include information about the possible strategies and the knowledge base required to do the task. It is needless to say that a theoretical model is needed for each task that relates skill performance to task performance. Here "thinking aloud protocols" or "prompted protocols" might provide better insights in how people really perform the task (process methodologies).

Factor-analytic models can only be useful in an exploratory or confirmatory way. They are regarded indirect methods to explain the phenomena. A better way according to Dr. Frederiksen is the analysis of protocols by experts.

He mentions the work of Robert Sternberg and Andy Rose as prominent examples with regard to the purpose and the intentions of the present project.

Dr. Daniel Gopher
Technion Haifa - Israel

Professor Gopher's main interests are in the area of general performance research. His orientation is both basic and applied. He uses a number of performance paradigms, especially dual tasks.

He believes that usefulness for theoretical and for practical purposes is not separable. For the domain of attention he regards focused-divided attention tasks (e.g. dichotic listening, dual task situations) as theoretically useful. From his point of view good generalizability does not necessarily require complex tests: "better a battery of simple tasks than a few complex tasks". Important are "tasks to get learning traces because the rate of progress would be a better predictor than performance level".

Therefor apart from speed and accuracy he considers as other useful performance parameters (a) the "rate of progress" (slope), (b) control over performance outcome e.g. to introduce consistent variabilities (by changing properties) or to stay in a certain window (single task), (c) transfer capabilities, and (d) ability to maintain performance constant when the level of difficulty varies.

As a main reason for the low generalizability to real life tasks he assumes the high variability of prediction criteria. Low generalizability could be avoided if (a) the prediction criteria would be worked out with people in the field and if the same work on statistics would be done, or concerning prediction procedures if (b) a combination of regression and cut-off methodology would be applied, (c) steps in the criteria would be developed instead of applying discriminant function analysis, and (d) the outcome of different or approximating procedures would be compared.

Breaking down real life tasks into components is considered as sensible, the development of a standard-battery as possible. Concerning the to be included tests he refers to publications of Wickens and himself.

He explicitly rejects factor-analytic approaches, because the meaning of factors always remains obscure. He recommends empirical testing against the criteria instead.

He is optimistic with regard to the possibility to develop a sufficiently broad battery. He believes that success depends on the criteria. There are no skill cateqories or classifications he could recommend in advance, because they depend on the definition of criteria.

Dr. Frederick Hegge
Walter Reed Army Insitute of Research, Washington, DC USA

The interview with Dr. Hegge was conducted to get an overview of the activities of the Walter Reed Army Institute of Research with regard to the development of a standardised task battery. Dr. Hegge gave an extensive report of the present activities that is summarized below.

Dr. Hegge stated that the Army battery was developed in the general context of medical or chemical defense - especially with regard to the effects of certain psychoactive drugs on the performance level. After an extensive drug screening and testing program they are now looking for the behavioral component of psychic drugs. This is achieved by 48 projects in 23 different laboratories. The research has involved the following stages:

(1) Level 1 focuses on the drug dose setting in the behavioral laboratory. This was achieved by means of a standardized task battery which included
(a) a neurophysiological battery (including EEG-measures)
(b) a psychomotor test battery (including measurements of microtremor and tracking)
(c) a neuropsychological battery

Here Dr. Hegge reports attempts to establish a computerized standardized neuropsychological battery ("standardized" means "agreeing to do the same thing") as a first step in standardization. That provides a foundation for development of normative systems. They are developing an "engineering" system rather than a "research" system.

Information about the drug dose came also from the Animal Behavior Group that investigates performance in stressful situations that cannot be done with humans.

(2) On level 2 the drug effects on performance are studied with human subjects in a residential screening facility. The major instrument to assess performance effects are the "Unified Tri Service Cognitive Performance Battery" (UTSCPB), a physical performance test battery, and a scale for the subjective assessment effects for mood and activation.

(3) Level 3 explores the effects of environmental and situational stressors like sustained attention and sleep deprivation.

In general, levels 1 - 3 aim at the biological and functional substrate of behavior. It is intended to isolate the major biological resources that determine the performance level. These projects are aimed at establishing a comprehensive descriptive data base about human performance.

Another line of research involves the use of simulation programs of real-life tasks ("command and control"). The simulation programs are developed in close contact with people who perform the task in daily life ("one foot in the field").

Here a major emphasis is placed on the task analysis of real-life tasks. Task analyis is done theoretically and empirically. The basic question behind the task analyis is what psychic function is affected to what degree by the drug or the environmemtal stressor and where can that psychic function be found in the theoretical and empirical task analysis. Another point concerns "sequential network modeling" where complete weapon systems like the M60 tank and scenarios have been simulated on a microcomputer. These networks are developed in cooperation with people who do the task. The focus is on the time to perform, internal errors, error correction and military outcome.
The task analysis data base serves a risk identification function. The sequential network models of man/machine crew/machine systems provide risk quantification estimates.

According to Dr. Hegge it is of critical importance that the development of such a battery can only be successful if one switches continously between laboratory and field research.

Dr. G. Hitch
University of Manchester, U.K.

Dr. Hitch's primary interests are in the areas of human memory, arithmetical skills and man - computer interaction. The main research paradigms in his experimental research are concerned with traditional human memory tasks but also include dual task and arithmetical task techniques. With regard to theoretical issues he is very keen on converging operations on the basis of different tasks. The probability of task-specific artefacts is high when relying on one simple paradigm only. He also aims at using tasks that can be well described in component aspects.

With respect to real life applications Dr. Hitch is aware of a gap between memory paradigms and "memory-in-real-life". Arithmetic tests have greater ecological validity as nad Bartlett's type of approach - but on the other hand there is a real problem of generalization in more complex memory tasks. Speed and accuracy, and measures derived thereof, will remain predominant in behavioral research. In addition, verbal protocols (e.g. thinking aloud), and more detailed analyses of types of errors and judgments could open interesting methodological avenues.

In order to improve the predictability of real life performance on the basis of laboratory tests, Dr. Hitch sees a clear need of new paradigms that should be as close as to what is found in real life (simulation). Alternatively laboratory tests should be developed that enable the measurement of basic cognitive capacities. Then, the real life task should be analysed in terms of these basic processes (Card, Newell and Moran - Human Computer Interaction). Whether this approach is successful depends on the nature of task organisation. If the components all mutually interact to constitute a new whole, one cannot expect basic components to be valid predictors of performance in the real life task.

Yet, Dr. Hitch is of the opinion that a battery is feasible. It should include tests of perception, attention, memory and motor control. Furthermore knowledge based skills (e.g. reading, arithmetic) and tests of the knowledge base itself (psycholinguistic skills, reasoning, spatial abilities) should be included. Dr. Hitch worries about arbitrary tests as found in the factor-analytic approach. Tests should have theoretical models underlying them.

The breadth of the battery depends on the extent that task specific knowledge plays a crucial role in performing the real life tasks. If this is generally important, the value of using performance in component tasks, as predictors of the real life skill, is bound to be limited. If not, a small battery is most promising.

Dr. Earl Hunt
University of Washington

Dr. Hunt gave some comments on related projects and scientific efforts in the same direction as our project:

(1) battery from Brooks Air Force Base (Ray Crystal), which he regards technically o.k., but training effects have not been taken into account

(2) battery from Army Research Institute (Wing) which is mainly a psychomotor battery

(3) battery from Bob Kennedy (comment: "psychometric tour de force")

(4) battery from Jim Pellegrino and Earl Hunt (available from March 1986) that is primarily concerned with coordination of motion including timing aspects.

He also mentioned an approach from the Educational Testing Service (ETS) that would be available in the spring of 1986.

Factor-analytic approaches are regarded as serious, but the methodology is a little bit out of date. A factorial design should be preferred.

Dr. John Ionides
University of Michigan, Department of Psychology

Dr. Ionides has worked in the areas of cognitive psychology and perception . His present research interests focus on scene perception. His primary research interests are basic.

In his research he has employed mainly reaction time as a dependent measure but also discrimination judgements where accuracy was the dependent variable. He believes that reaction time tasks have a fair degree of validity in human performance research especially with regard to those real-life tasks that have a speed component.

However, he mentions three different metrics that might be useful in human performance research. First, similarity judgements might provide some insights in the internal representions that a person has of a set of stimuli. Multidimensional scaling techniques are a powerful method here. Second, a variant of accuracy should be examined more closely. In general, the nature of errors that people make is neglected in looking only at error percentages. The nature of errors, however, may reveal a lot more about the structure of the cognitive system and about the strategical aspect of behavior. (e.g. separating between intrusion-, ommission-, and confusion-errors). Third, protocol analysis is generally an "awful" method, but maybe useful as a heuristic tool to generate hypotheses about process characteristics of human performance. It should never be used as a dependent variable, however.

The reasons for the low validity of performance tests with regard to the prediction of real life performance can be attributed to the fact that real-life performance is much more subject to strategical influences of how the subject organizes his/her performance. It is part of the intended nature of the laboratory task to deprive subjects of their strategical freedom. Quite the opposite holds for real-life performance where within certain constraints the subject has multiple strategies how to do the task. Depending on the strategy that the subject chooses the single component gets more or less important within the whole process of task performance.

A possible way to improve the generalizability of laboratory tasks is to give up the restriction of at the most two response alternatives and thus approach the strategical freedom of a real-life task. It is self-evident that in this case process hypotheses on the various response alternatives should exist.

Dr. Ionides is positive towards the idea of being able to break down a real-life task into distinct components. He mentions the work of Bob Kennedy as an example.

However, he is a little bit worried about finding a finite number of tasks that capture the broad domain of skills usually found in real-life tasks. It might be possible for a limited domain of natural tasks (e.g. complex visual processing) which does not necessarily have to be trivial.

He does not know about any skills categories but mentions the work of Robert Sternberg ("Beyond I.Q.") where intelligent behavior is conceived as being based on a number of underlying cognitve abilities and the work of David Buss (Psych. Review, 1984).

Dr. D. Jennings
University of Pittsburgh, Pa, USA

Dr. Jenning's primary interest is in basic research in cognitive psychophysiology. His research has centered around relations between performance tasks (RT, recall, recognition) and physiological variables. In his basic research he feels that the tasks should be as simple as possible to permit tests of theoretical issues. With respect to applied questions his opinion is that more complex, simulation type techniques might be optimal. He is clearly aware of a gap in this respect. With regard to long-term applied aims he would try to arrive at generalized variables--permitting general rather than highly specific predictions. He suggests that laboratory/ theoretical work is necessary to analyze a task into its components and the variables influencing those components. This information should not be expected to be directly relevant to field/ real life performance. Performance in field settings will be determined by a large number of factors not present in laboratory. The commander/ manager in the field must relate known variables affecting the task (i.e. lab knowledge) to existing conditions (i.e. practical knowledge, intuition) and predict performance in that setting. He believes that this is the only practical--i.e. cost effective--way of using performance research. The traditional measures of speed and accuracy--or some more sophisticated derivate--seem to be the only feasible measures, of course apart from physiological concommitant measures.

The usually observed low validity of individual performance tests with regard to real life tasks might be at least partly due to differences in context and practice. It is the assembly of component skills which may occur uniquely in real life. In addition the components, as well as the way of assembling, are highly practiced. Laboratory tasks may never reach a high level of generalization if the capability of assembling component skills is not considered. Varying the learning set to identify assembly rules may be a promising approach.

Real life tasks may be broken down in their components in order to enable some degree of comparison between tasks. It remains to be seen whether this has validity. A standardized battery may be constructed with regard to components. Yet, since the assembly element is not considererd, the direct applied value should not be oversold.

Possible tests of a battery could be a) choice-RT, b) tracking, c) STM/LTM, d) dual task capacity, and e) ways of combining such elementary tasks.

Dr. Jennings is not impressed with factor-analytic correlational approaches because of the atheoretical haphazard nature. His final comments concern the breadth of a possible battery: It should certainly not be too broad, since too many subskills would be involved. Limiting to a couple of skills - such as flying, car driving - would be optimal.

Dr. Daniel Kahnemann
Dept. of Psychology, University of British Columbia, Vancouver, Canada

---

Dr. Kahnemann first comments on the intentions of the present project. According to him it would be a desirable venture although he feels that some people would refuse to adopt a positive outcome. However, if the project succeeds in providing some standard versions of laboratory tasks he would clearly regard this an an advantage and a starting point for future research.

In his own research he has been employing mostly choice-reaction-time tasks, but also detection tasks with detection probability as the main dependent variable, visual search tasks, and visual memory tasks.

He regards reaction time and percentage correct the principal metrics in human performance research although some measures that can be derived from these two have turned out to be useful (e.g. slope measures). In that context the relative position of the individual on the "speed accuracy dimension" provides important information about the more strategical aspects of behavior.

He is not surprised at the low validity of laboratory task with regard to the prediction of real life performance. Laboratory tasks are usually picked up at a very low level of practice whereas real-life tasks are usually highly practiced. That is the key to the low validity. A test in a test battery is always limited in time (usually not longer than 30 minutes). It is hopeless to believe that a preliminary test of a single skill should have predictive validity for a highly practiced complex task where this skill interacts with numerous other skills and that interaction is directed by different strategical "supervisors". If a test is incorporated in a battery that is supposed to predict complex and highly practiced human performance then this test must be predictive for the final performance level. According to Dr. Kahnemann this is an absolute "must" for each test being incorporated in a battery with such an aim.

Therefore the generalizability of laboratory tasks can only be improved if these conditions are met. It seems doubtful whether the number of available laboratory tasks are useful here. He feels that the standard laboratory paradigms are worn out a little bit. Researchers should consider new paradigms.

With regard to the question to what degree a real life task can be broken down into components Dr. Kahnemann sees no general answer. A breaking down seems possible to him for a number of small tasks whose structural and functional demands on the cognitive system are well known. These should also be tasks where the probability of a failure is very rare.

He mentions the work of John Duncan at the Applied Psychology ... in Cambridge (UK) as related to the topic of the project.

Dr. Steve Keele
Department of Psychology, University of Oregon

Dr. Keele's primary research interests are attention and motor processes. In his experimental research he has employed motor timing tasks with intertap variability as a principal dependent measure. He has also dealt with force control measures, time-sharing paradigms and measures of attentional flexibility.

Additional important measures in human performance research are measures of vigilance decrements in sustained attention tasks and measures that are designed to depict interindividual variability in strategies of task performance. A good example are the current approaches in reading research that try to break up the reading process in an analytic way (Hunt et.al.) to predict reading comprehension and reading errors. Also motor timing has turned out to be a variable that differentiates between different levels of attentional flexibility (see also the work of Navon & Gopher, 1979).

Usually real life tasks are complex tasks in the sense that they involve a lot of components that are likely to interact with each other. A real-life task can be carried out by using different strategies. Furthermore these tasks are usually highly practiced. All these features do not apply to laboratory tasks. These tasks are never extensively practiced, they are designed to study single phenomena in an artificial context and therefore the strategical freedom of the subject is rather limited.

Being able to predict the performance in a real-life task requires a deep understanding and a thorough analysis of the processes and interactions involved in that task. Even for a rather simple task this can require quite a few years of investigation.

Dr. Keele feels that a standard set of subtests might create some problems because the selection of subtests probably depends on the task that is investigated and possibly on the state of the subject (e.g. in "drug"-research). It will not be possible to assess the large variety of real-life tasks by means of a limited number of laboratory tests and still expect a good predictive validity.

His attitude towards factor-analytic approaches seems to be negative because it is not a process approach that can depict the interindivually different ways to perform a task.

Dr. Keele reports that the work of Harold Hawkins at the Office of Naval Research is related to the aims of this project.

Dr. Gordan Logan
Department of Psychology, Purdue University

Dr. Logan is working in the area of attention and performance. His present research interests focus on automaticity and the inhibition of thought and action. His primary research interest is basic.

In his own research he has employed choice-reaction-time tasks like the Sternberg-paradigm or the Stroop-paradigm, but also lexical decision tasks, category judgements, and visual search.

As particularly useful with respect to theoretical developments he regards any task that is well understood. Even an old task looked at from different viewpoints may be theoretically fruitful (e.g. the repetition effect). Laboratory tasks have share some features with real-life tasks but these features (e.g structure of the display) may be rather different. the Stroop-task e.g. ist not considered to be ecologically valid.

Dr. Logan feels that reaction time and error percentage are still the principal metrics in human performance research. However, ratings of workload, evoked potential analysis and so-called "rate measures" (bits/second) which put together speed and accuracy, are prominent alternatives to the standard metrics. He generally believes that most of the metrics are derived either from speed or from accuracy or both. The interference effects observed in dual task experiments also represent an alternative to the classic metrics.

With respect to the low validity of laboratory task for the prediction of real-life performance Dr. Logan states that the procedures of experimental tasks are not similar to real world tasks. E.g. there are generally no circular arrays in visual search tasks under real-life conditions. This may heavily influence the top-down strategy of visual search that the subject selects. Furthermore the pronounced interindividual differences even in simple tasks and the various strategies of performance are not considered in laboratory research. In general there is an ignorance of strategies and different abilities. Since the deprivation of strategies is part of the philosophy of experimental design, laboratory tasks can never reach a satisfying predictive validity with regard to real life tasks.

A way out of this might be the making-up of new tasks that are closer to the real-world tasks. A changing of the parameters of already existing ones may be an alternative. Dr. Logan feels that the present laboratory tasks are not designed to have a high predictive value. The best solution would be to make an analogue of the real-world task. In that case it must be known what the basic abilities are that are relevant in performing the task. In this computational approach (find out the basic abilities and combine them) it is absolutely mandatory to carry out a very detailed task analysis with regard to the functional and structural resources involved.

It heavily depends on the extent to which the structural and functional components of task performance interact whether a real life task can be broken down into components. Depending on the degree of interaction the validity will increase or decrease. He sees some possibility to achieve the aims of this project for small tasks whose structure and demands are well known. Developing a broad enough battery with a finite set of subtests for more complex tasks occurs to him a big piece of work.

Dr. Dominic Massaro
University of California at Santa Cruz

With regard to methodology Dr. Massaro has most frequently employed identification judgements - mostly in connection with factorial designs. He mainly used reaction time as a dependent measure as well as the percentage correct in these judgements. The experimental settings usually required "yes/no"-judgements, but also continous judgements in some cases.

Dr. Massaro regards basically every method as theoretically useful as long as a number of variables can be manipulated and valid conclusions can be drawn from these manipulations that lead to advances in theory building. Rating scales are thought of as particularly useful in the assessment of skills.

The low validity is based on the fact that there is only a partial overlap between processes involved in a real-life task and a laboratory task. If one generally succeeds to produce a high degree of overlap one can expect a better validity. In the end this should result in the laboratory simulation of complex real-life tasks that come close to the real situation.

With regard to factor-analytic approaches Dr. Massaro emphasizes that these methods only can have a heuristic value. A better way to explore the architecture of the cognitive system is by model building and testing.

Dr. Merrill Noble
Department of Psychology, Penn State University

Dr. Noble's area of specialization is research on human performance. His basic research interests are attention and motor control. His research interests are more applied.

He has been mainly involved in laboratory type of research and has employed almost any experimental method, but mostly reaction time measures in the additive factor tradition to infer stages of processing in serial choice reaction time tasks.

Thus he regards reaction time methods as particularly useful with regard to theoretical developments. On the contrary, he belives that reaction time methods are not very useful with regard to more applied situations because reaction time methods only have predictive value when the subjects in a real life task are under a comparable time pressure which very rarely occurs.

Other possible metrics in human performance research are information rate measures or subjective measures (rating scales). Dr. Noble is very reluctant towards physiological measures compared to performance measures ("they don't tell me something I don`t know").

The low predictive validity of lab tasks with regard to real-life performance is caused by the fact that real life situations involve a lot more operations. Unfortunately very few things are known about real-life tasks so that the components are not fully known. Almost nothing is known about the interactions of component processes in real-life tasks. Dr. Noble recommends going back and forth between theoretical studies and applied studies - i.e. between laboratory studies and the investigation of real-life tasks. This should also include simulation studies. In fact, Dr. Noble regards simulation studies as the most important way to achieve a satisfying predictive validity.

Basically it seems possible to break down a real life task into components but one should be aware that the more you decompose the less predictive validity can be expected. Therefore it seems necessary to think about the general philosophy of the project with regard to this question. However, for some tasks it seems conceivable that for some tasks (e.g. visual search) a reasonable validity can be expected.

Dr. Noble states that in any case it seems necessary to specify the number and kind of tests dependent on the specific real-life task under question. A general battery that covers the large variety of human behavior seems not possible to him at the moment.

Factor-analytic approaches are not to be considered as major sources of information in these kinds of problems.

Dr. R. Näätänen
Department of Psychology, University of Helsinki, Finland.

Dr. Näätänen's primary research interests are concerned with orienting responses and mechanisms of attention - both from the physiological and the behavioral point of view. A combination of physiological - mainly evoked responses - and performance tests is characteristic for his research. The behavioral tasks included disrimination thresholds and simple RT tests, but also simulation of risky situations. His main emphasis is on basic research with regard to theory, he considers the study of evoked potentials as particularly useful, since it follows the actual process in the brain and suggests which areas are activated by certain stimulation and performance. He does not see a basic difference in theoretical and real life research techniques. Methods such as the evoked response should be further developed so as to deliver relevant information about real life.

Apart from the traditional speed and accuracy measures, Dr. Näätänen suggests measurement of safety margins (risk taking), and related decision making, as well as endurance measures. He agrees that most laboratory tests have a low validity and argues that with the common speed and accuracy measures in simple tasks, one fails to tap central decision elements, that are so characteristic for real life tasks. For instance, a main problem with the Häkkinen - battery on driving—when applied to private drivers—is that it has too much emphasis on perceptual-motor skills. (With bus drivers and others performing in not self-paced tasks the Häkkinen battery works very well.) The improvement of real-life prediction requires that judgmental aspects rather than perceptual - motor overload are taken into account (cf. Näätänen, R., & Summala, H. (1976). Traffic accidents. Elsevier, NL: North-Holland).

Breaking down tasks in components may be sometimes possible - e.g. traffic - but is certainly not easy; various features of performance are hidden and can only be seen after prolonged work. A standardized task battery may work for limited sets of real life tasks. Dr. Näätänen doubts whether such a battery will have general value. If constructed, a battery could include some of the better researched tasks - e.g. memory search, dichotic listening etc. - but one should be careful to trust them too much. Factor analytic approaches are no good entry; according to Dr. Näätänen they will not work.

Dr. Raja Parasuraman
Catholic University - Washington, DC

Dr. Parasuraman's main scientific and research interests are in the area of attention and vigilance. He considers his own research to be basic as well as applied. He stresses that his comments on the feasibility of a standardized battery are limited to this special area. As main research paradigms he has used attention and vigilance tasks, especially discrimination, choice reaction time, and dichotic listening tasks. He considers these tasks to be particularly useful for theoretical purposes, but useless concerning generalizability to real life performance. He assumes that speed and accuracy are the essential performance measures. Reasons for the poor generalizability he sees in constancy of laboratory situations, large inter-individual variance of performance levels in the field, and low correlations of laboratory tests between eachother. Generalizability might be improved by paying more attention to inter-individual performance differences and controll outside the lab.

Dr. Parasuraman beliefs that real life tasks can be broken down into components, but there are other factors in reality which must be taken into consideration, - this can easily be shown for driving performance for example.

He thinks that the development of a standard-battery of performance tests is possible, but rather difficult. The starting point should be the development of an information processing model. He regards factor-analytic approaches as bad, because the mathematical procedure does not take dynamic processes into consideration.

Of particular difficulty assumes Dr. Parasuraman the development of a test-battery, which is sufficiently broad to cover all the most important real life skills. He suggests that a successful battery might be possible only for limited areas of skills like car driving etc.

Concerning vigilance tasks he regards a classification as possible, which takes different strategies into consideration.

Dr. M. I. Posner
University of Oregon, Eugene, Oregon U.S.A.

Dr. Posner's primary interests are in the basic aspects of human attention and performance, viewed from the behavioral as well as from the neuropsychological side. His main paradigms are chronometric, and with regard to the analysis of performance, he aims at using as simple tasks as possible. In applied research the situation is different. Dr. Posner does not feel that a real life task can be easily broken down in components. Yet he feels that the Robert Sternberg approach may have future.

Apart from the traditional speed and accuracy measures, he mentions (1) learning rate, (2) protocol analysis and (3) eye movement protocols, as valuable tools for behavioral analysis. Whether one is capable of predicting real life from these measures is doubtful, although they should provide the basic insights and building stones to recommend about actual tasks. The major problems in direct correlational prediction are motivational and organisational, in that the social context is absent in the experiment. If you can free the real life task from the social context one can do quite good with elementary tasks. He does not agree with Neisser's "Cognition and Reality". If there is a good task analysis then one can find basic components. One line of evidence is the expert system approach, but some of the well investigated laboratory tasks should do well also.

Hence, Dr. Posner considers the construction of a task battery as feasible, although probably better for sensory-motor tasks than for more abstract command and control. The construction should start with specifying some major cognitive systems, such as: object recognition, several varieties of attention, motor control, lexical access in language. The next step is to know which tasks refer to which real life tasks. Here task analysis - perhaps also through writing an expert system - is required. Mapping the components to the task is the final step. The factor analytic approach is not favored by Dr. Posner: It is too atheoretical. The Sternberg approach is preferred. In this way a broadly predictive battery should be possible - unless emotional aspects interfere too much. But Dr. Posner feels that progress is also possible in that direction, for example by studying achievement motivation.

Dr. Walter Schneider
Department of Psychology, University of Pittsburgh

Dr. Schneider is working in the areas of attention and skill acquisition, especially with regard to the effects of practice on automization of certain aspects of behaviour. In the area of human performance research he has dealt with air traffic control, EEG measures, and skill acquisition in electronical troubleshooting.

Particularly useful with regard to theoretical developments he regards those methods that focus on the representation of knowledge and change of knowledge. The "dual task paradigm" also plays an important role. Except speed and accuracy Dr. Schneider names physiological indices (e.g. the P300-component of the EEG) and "time on task" (TOT) as important measures of performance.

According to Dr. Schneider the low validity of performance tests in predicting real-life performance must be expected because a) real life performance usually is highly practiced, b) real life performance is heterogenous with respect to the various components involved, and c) real life performance generally is no good predictor for other real life tasks. This means that in general test performance is the "psychology of the first 30 minutes" of a person performing a task. For him the reasons for the low predictive validity lie within the integrating effects of extended practice on a task.

One possibility, however, to improve the generalizability of laboratory tasks is to make them gross measures in the sense that they should not be restricted to measure an isolated process. The last consequence of this idea means that the real life task should be simulated in the lab. In that case a better validity can be expected.

The degree to which a real life task can be broken down into components depends on the ability to identify the appropriate cognitive units behind the units of a task analysis. It is decisive here not to carry out a task analysis in the traditional sense but a cognitive component analysis. Another important point here is that this analysis should also take into account the effects of practice. Every model of human performance that does not include predictions on what changes during practice cannot be expected to be an adequate model of human behaviour.

Dr. Schneider is not thrilled by the factor-analytic approaches since these explain only a relatively small percentage of the performance data. Above all the basic assumption acoording to which the human mind is a linear system seems at least questionable. There is no way to believe that human cognition is linear.

He does not deny the feasability of developing a standardized battery of performance tests but this program depends heavily upon finding (new) tests that have at least some predictive value with regard to real life tasks. As examples he mentions an approach by Alan Baddeley and his own work together with Phil Ackerman.

Dr. Wolfgang Schönpflug
Dept. of Psychology, Freie Univerität Berlin, F.R. of Germany

---

Dr. Schönpflug's main interests are in the area of general and experimental psychology. His current research interests concern action theory and human factors. He considers them to be basic and applied. He has analyzed behavior in complex situations, e.g. simulation of work on computer displays, administrative and planning work. Apart from conventional performance measures he is particularly interested in efficiency (ratio performance/effort), strategy development, and rate of progress.

As a main reason of the poor generalizability he considers the lack of consistency concerning number and organization of the components of a task. He does not believe in the success of a general task battery, but more specific batteries for example for places of work requiring sensu-motor coordination or places in administration or management might have better chances.

He would be more positive toward factor-analytic approaches, if they would be applied with more sensibility.

Generally he prefers simulation of complex real life situations to the development of laboratory task batteries.

Dr. Gordon Shulman
Department of Psychology, Penn State University

Dr. Shulman's areas of specialization are attention and spatial vision. His present research interests are spatial attention and spatial frequency channels. His primary research interests are basic.

He mainly has employed reaction time measures in a cueing paradigm (effects of advance information). He also used probe methods, dual task paradigms and measures of contrast sensitivity. According to him the results of probe methods can be generalizable - especially in a dual-task context.

Other measures except reaction time and error percentage are physiological measures like latency and amplitude of the P300 component in dual-task contexts and pupil changes.

The main reason for the low validity is that the experimental psychologist has designed laboratory tasks to isolate a special process that he likes to study. So basically all the other context variables are considered to be contaminating. Quite the opposite is true in performance assessment in real life tasks. Here a phenomenon cannot be studied in isolation. However, some laboratory tasks may have ecological validity. The distribution of visual attention (visual search tasks) in the laboratory and searching for a friend in the crowd are possibly governed by the same processes. Visual search seems to be a rare example where a laboratory task involves roughly the same processes as a real-life task.

Dr. Shulman thinks that the attempts to break down a real life task into components have been very successful in the past. As an example he names the dichotic listening task of Kahnemann. He is very sceptical towards these approaches. He feels it is better to simulate the task to get a better validity. For this project's approach he sees no chance at the moment.

If, however, such a battery is planned this battery should include a perceptual measure like contrast sensitivity, a measure of short-term-memory like a digit-span task or a Sternberg-like task, some test of the ability for visual-motor coordination like a tracking task. Furthermore a test of the speed of retrieving linguistic information and a task to manipulate spatial information should be included. However, according to Dr. Shulman, it seems not possible to come up with a finite number of tests that cover the most relvant aspects of real life-performance. It might be possible for a well described real-life task. Thus every real-life task would require a different battery depending on quality and quantity of cognitive processes involved.

Dr. Shulman reports the following scientific approaches that are related to the aim of the project:

- the work of Diane Damos with regard to time sharing ability (e.g. in busdrivers)

- the work of Earl Hunt with regaru to information processing correlates of reading
- the work of Meredith Danaman on the role of working-memory in a number of information processing tasks
- the work of Harold Hawkins
- the work of Edwin A. Fleishman
- the work of Alan Baddeley of the MRC Applied Psychology Unit with regard to the effects of environmental stressors like carbon dioxide and heat on human performance.

Dr. Robert Sternberg
Dept. of Psychology, Yale University, New Haven, Connecticut

Due to time pressure the interview with Dr. Sternberg was very short. However, he assured that most of his views towards an issue like this had been laid down in his book "Beyond I.Q.". After having read the information sheet he mentioned the work of Andy Rose as related to the aims of the project. Furthermore he strongly advised not to start with the standard laboratory tasks but with the thorough analysis of real-life tasks. Only if the structure of a real-life task is extensively known one could think of the appropriate laboratory tasks to measure components of the real-life tasks. The available laboratory tasks might not be very good candidates with regard to our programm since they were developed for very different reasons.

He also thinks that reaction time and error percentage are the prominent measures in human performance research, but decision probalities or probabilities estimates for a certain event may be some alternatives that are not based on these measures but tap different psychological processes.

The reasons for the low validity of performance tests lie within the context-reduced nature of laboratory tasks which are designed to study an isolated phenomenon under quite artificial conditions. Therefore laboratory tasks cannot be taken without some "grains of salt" to predict real-life performance.

Dr. Sternberg's attitude towards factor-analytic approaches is that they serve a heuristic purpose but that a modelling approach should be preferred in the assessment and identification of basic information processing components.

With regard to other skill categories or classification of skills he refers again to his book "Beyond I.Q." and to the work of Nancy Anderson at the University of Maryland.

Dr. Max Vercruyssen
USC, Los Angeles, USA

Dr. Vercruyssen is especially interested in basic and applied performance diagnosis. He has got experiences with a battery of 30 tasks. In principle he applies a multi-method approach (physiological indices, subjective and performance measures) and/or a multi-stressor approach. He selects tasks step by step, looking for sensitive performance measures first and applying additive factor methods to these measures second.

Concerning generalizability he believes that an approach closely related to back-to-back experiments is needed. As a main reason for the poor generalizability he assumes the difficulty to have the whole bandwith of the real world. Generalizability might be improved by clear representation of the components of real life tasks and back-to-back experiments.

Apart from conventional performance parameters he considers bias parameters (S/N ratio) and state parameters (subjective and physiological) to be important.

Breaking down of real life tasks into components has to be considered as difficult.

Tasks with a good tradition should be selected, but beyond that also many others.

He regards the Fleishman approach as a very good approach which should be revived.

One battery to cover most real life skills is considered as unrealistic. Dr. Vercruyssen rather suggests multiple batteries.

Dr. Christopher Wickens
Dept. of Psychology, University of Illinois, Champaign, Il. USA

Dr. Wickens' main interests are in the area of Aviation and Engeeering Psychology. His current interests concern the whole range of human performance theory, including attention, manual control, decision making, work load, and automatition. He considers his interests to be basic as well as applied. He has been using a long list of experimental methods and paradigms, including dual task, Sternberg task, tracking (stable and unstable critical), maze tracing, embedded figures, dichotic listening, evoked potentials (auditory and visual), and the Brooks-Matrix-Test. He considers dual task, Sternberg task, stable tracking, embedded figure test for measuring cognitive style, and the Brooks-Matrix-Test to be theoretically relevant. Relevant for real life skills are in his opinion stable and unstable tracking, less important are maze tracing, dichotic listening, and evoked potentials. Concerning real life skills he recommends to use nonstandardized complex tasks, like simulation of process control, trouble shooting (diagnosis), or aircraft control simulation.

Apart from speed and accuracy measures he recommends performance parameters like bias (signal detection indices, tracking gain), style (speed/accuracy trade off in reaction time tasks), and resource indicators (e.g. P300-amplitude).

He suggests that the poor generalizability of real life tasks may be due to the high variance of motivation and the fact that in contrast to lab tasks real life tasks are multitask situations. A possibility to improve generalizability might be to implement tests in multitask situations (at least dual task situations). According to his own experience single tasks are capable of explaining only 20 -40% of the variance of a simulation test.

Dr. Wickens considers the development of a standard-battery as positive. It should include tasks of the following kind: Sternberg-type tasks, critical unstable tracking, memory tests (running memory, digit span, spatial memory), dual task (Sternberg + crit. tracking), planning and scheduling test (Tolga and Sheridan), mental rotation.

Concerning the factor-analytic aproach he has not got a decided oppinion.

He sceptically views the possibility to contruct a battery broad enough to cover most real life skills: "the tests may cover them, but only 20 - 30% of variance".

Dr. Gery d'Ydevalle
Dept. of Psychology, University of Leuven, Belgium

The areas of specialization of Dr.d'Ydevalle are cognition and motivation. His primary research interest is basic. He mainly employed the Posner comparison task as an experimental paradigm or what he calls "free movement situation". He never used tachistoscopic presentation of stimuli.

He is very reluctant towards the possibility of disentangling some basic structures of human mind. On this argument he bases his criticism of the factor-analytic approaches. The rich structure of cognition probably cannot be reduced to a very few dimensions. Even if one succeeds in extracting some of the basic components there would be no way to assess the multitude of possible interactions between these components. Endeavours to assess some basic components, however, might be fruitful.

According to Dr. d'Ydevalle the limited repertoire of performance measures should be augmented by measures that focus on the strategical aspects of behavior. Here he mentions measures derived from assessment of eye-movements as a possible candidate.

With reoard to the question whether a real-life task could be broken down into distinct components that can be assessed separately he sees some possibilities that this might be achieved. However, some major problems arise when these distinct components have been identified only conceptually and no measurement procedures exist to assess them in a reliable and valid way.

5.3. REVIEW OF THE TASK BATTERIES

This chapter contains a preliminary inventory of the most commonly used test batteries in human performance research and stress research. The inventory will not necessarily be exhaustive but covers the most relevant developments in this area. The sequence of batteries is arbitrary.

We have reviewed each battery according to the following general scheme:

(a) General description of the battery:
- authors
- title
- source
- reported original purpose
- reported criteria for the selection of subtests
- reported validation procedures
- reported theoretical background for the whole battery

(b) Specific description of subtests
- main references
- theoretical background/ performance domain
- stimulus materials
- procedure
- administration time
- scoring and norms

List of the reviewed task batteries
(1) BAT - Basic Attributes Test
(2) BBN - Test Battery
(3) CTS - Criterion Task Set
(4) IPT - Information Processing Tasks
(5) PAB - Performance Ability Test
(6) TTP - Taskomat
(7) HAK - Test Battery

5.3.1. BAT - BASIC ATTRIBUTES TEST

(a) General description of the battery

- authors:
Kantor et al.

- title:
BAT (Basic Attributes Test)

- source:
interview with Dr Kantor, Brooks AFB, USA, November 1985

- reported original purpose:
This attempt to develop a standardised task battery serves the more limited aim of constructing a new pilot selection battery. Although most of the tests are performance tasks, the battery also includes personality questionaires. Hence it is referred to as covering "attributes", rather than "abilities" or "mental functions".
During the interview with the senior investigator of this project, Dr Kantor, Brooks AFB, USA, it was made clear that, although the BAT does not pretend to be a general standardised battery, the general idea is still that most relevant cognitive and perceptual-motor functions are adequately covered.

- reported criteria for the selection of subtests:
included feasibility, interest of the test-taker, independence from other tests, construct validity and minimal dependence on verbal material.

- reported validation procedures:
At the time of the interview the BAT has not yet in operational use, but the determination of the predictive value of the separate tests with regard tc (fighter) pilot success in training were currently underway.

- reported theoretical background for whole battery:
The BAT leans heavily on the results of human performance research of the last few decades, in that most of the tests have a firm background in basic research. On the other hand, the aim of a selection battery obviously requires correlational studies stressing individual differences in the test as well as in the real task performance criteria.
The ultimate choice of the type of tests, included in the BAT was also stimulated by the wide range of factorial studies of FLEISHMAN and coworkers as summarised in AFHRL Techn.Rep. 80-27.

(b) Specific description of the subtests

All tests of the BAT make use of a regular VDU display. Two joysticks - one to the left and one to the right of the subject's positions - and a 4X4 matrix keyboard, located in between the joysticks, serve as controls.
In the present operational testing of this battery, the total testing time, including practice and instruction, lasts four hours. This means that no

skilled performance can be expected on any task.
Main references for all subtests are to be found in Imhoff & Levine (1981).

*** BAT 1: PERCEPTUAL SPEED ***

- theoretical background/ performance domain:
Fitts law (Fitts & Peterson, 1964); motor programming of a sequence of responses (Sternberg, Kroll & Wright, 1978).

- procedure:
Four digits are simultanuously presented on the VDU in a horizontal row. The subject responds by releasing a homekey—located underneath the matrix --and pressing the corresponding succession of keys on the keyboard. There are two or three practice trials.

- administration time:
approx. 6 minutes

- scores and norms:
Although errors are recorded , the main emphasis is on reaction time (time to release the homekey), on movement time (time betwen starting the movement and pressing the first key), and on interresponse times.

*** BAT 2: DOTS ESTIMATION ***

- theoretical background/ performance domain:
subitizing; estimation of number and density; psychophysical scaling (Stevens).

- procedure:
The VDU display is divided into two equal parts by a vertical dividing line. In both halves a number of dots is presented which always differ by one dot. The dots appear in random positions at either half of the screen. The subject indicates which half contains more dots by pressing the corresponding left or a right key. The total number of dots is varied, so as to obtain a function relating response time (and errors) to difficulty of discrimination.

- administration time:
approx. 5 minutes

- scoring and norms:
reaction time.

*** BAT 3: TIME SHARING ***

- theoretical background/ performance domain:
adaptive compensatory tracking (Poulton, 1974 & 1981); dual task performance (Wickens, 1984).

- procedure:
On the VDU, a schematic front of an airplane is displayed together with a gunsite. The task consists of compensatory tracking - i.e. keeping the gunsite aligned with the plane - the difficulty of which is adaptive to performance (root mean square error) by varying the gain on the joystick. Subjects receive five 60 secs tracking trials, they receive further tracking trials in combination with visual digit cancellation. Each time a digit is presented on the screen which is replaced by a new digit when the appropriate key of the keyboard has been pressed.

- administration time:
approx. 30 minutes

*** BAT 4: ENCODING SPEED ***

- theoretical background/ performance domain:
simultaneous matching, same/ different responses (Posner, 1978).

- procedure:
Two letters are simultaneously presented, consisting either of capitals, normals or a combination. The subject's task is to carry out a same/ different response on the basis of physical identity or name identity in brief separate sessions .

- administration time:
approx. 5 minutes

*** BAT 5: MENTAL ROTATION ***

- theoretical background/ peformance domain:
mental rotation (Cooper & Shepard, 1973); imagery.

- procedure:
In this test the VDU is divided into two parts by a vertical line. In the left part a letter (F,G, or A) is presented for 2'', which is followed by a masking field. Then a rotation (60, 120, or 240 degree) is presented on the right part of the VDU, consisting either a plain rotation or a mirror image of the original letter (when rotated clockwise). The subject's task is to decide whether the rotation is plain or mirror imaged.

- administration time:
approx. 25 minutes

*** BAT 6: ITEM RECOGNITION ***

- theoretical background/ performance domain:
memory scanning (Sternberg, 1975).

- procedure:
A number of 1 - 6 digits is presented on the VDU in a horizontal row. After

presentation a probe is presented. Subjects are asked to indicate whether the probe was present or absent by way of a speeded response.

- administration time:
approx. 20 minutes

*** BAT 7: IMMEDIATE/ DELAYED MEMORY ***

- theoretical background/ performance domain:
running memory; continuing memory (Sanders & v.Borselen, 1965), keeping track of several things at once (Yntema & Schulman, 1967).

- procedure:
A series of digits is presented on the VDU. The subject's task is to react to each digit when the next one is presented by an adequate keypressing response. For example, a "2" may be presented follwed by a "3". During the presentation of the "3" the reaction to the "2" is given etc..

- administration time:
approx. 20 minutes

*** BAT 8: DECISION MAKING SPEED ***

- theoretical background/ performance domain:
(Sanders, 1980)

- procedure:
A traditional choice reaction test, in which one of the digits 0 - 9 is presented followed by a speeded reaction by releasing the homekey and pressing the adequate key from the matrix.

- administration time:
approx. 20 minutes

*** BAT 9: RISK TAKING (GAMBLING) ***

- theoretical background/ performance domain:
decision making, risk taking, maximizing profits (Edwards, 1966).

- procedure:
A 5X2 matrix of square boxes is presented on the VDU, containing the digits 1- 10 in the natural left to right and top-down order. Subjects are told that there is a "disaster" behind one of the boxes. They can open as many boxes as they wish and earn 10$ per box as long as they do not hit the "disaster". Hitting the disaster means that all earnings of that trial are lost.

- administration time:
approx. 15 minutes

***** BAT 10: EMBEDDED FIGURES *****

- theoretical background/ performance domain:

Gestaltpsychology, the forest and the trees in perception (Navon, 1977).

- procedure:

A target shape is shown on the VDU (e.g. a tilted rectangle). This followed by two complex figures, one at the left and one at the right of the VDU. The subject's task is to indicate by a left/right key press in which complex figure the target is embedded. A maximum of one minute is allowed during which period subjects usually come to a decision. Reaction time is the main measure, but accuracy is stressed in the instruction.

- administration time:

approx. 15 minutes

- scoring and norms:

reaction time

***** BAT 11: SELF CREDITING WORD KNOWLEDGE *****

- theoretical background/ performance domain:

measurement of meaning, semantic memory (Osgood, et al. 1957).

- procedure:

Lists of words are presented in succession. At the presentation of a word subjects indicate the meaning by multiple choice. There are 'easy', 'medium' and 'hard' lists, and, prior to the presentation of a list, subjects predict how well they will do.

- administration time:

approx. 5 minutes

- scoring and norms:

number of correct responses

***** BAT 12: ACTIVITIES INTEREST INVENTORY *****

- procedure:

On the VDU two possible activities are presented that are similar in nature, but one activity is slightly more risky than the other. For example: Swimming in a pool or swimming in the ocean. Subjects express their preference in each case.

- administration time:

approx. 10 minutes

***** BAT 13: TWO HAND COORDINATION AND COMPLEX COORDINATION *****

- theoretical background/ performance domain:
tracking; multihand coordination (Poulton, 1974).

- procedure:
There are two versions, namely two-hand pursuit tracking (by means of both joysticks operated by both hands) of an elliptical track, and two-hand compensatory tracking of a two-dimensionally moving target and a vertical rudder.

-administration time:
10 minutes per trial

- scoring and norms:
horizontal and/or vertical error from target

5.3.2. BBN - Test Battery

(a) General description of the battery:

- authors:
Pew, R.W., Rollins, A.M., Adams,M.J. & Gray,T.H.

- title:
Development of a Testbattery for Selection of Subjects for ASPT Experiments.

- source:
Bolt, Beranek & Newman Inc.
Report No. 3585
29 November 1977

(In the following text this battery will simply be called BBN)

- reported original purpose:
This test battery has been developed for selection of subjects for Advanced -Simulator-for Pilot-Training (ASPT) -experiments, esp. for matching subjects or to provide covariates for studies on success in pilot training.

- reported criteria for the selection of subtests:
(1) high potential validity for predicting success in pilot training;
(2) accumulated time for testbattery administration should not exceed two hours per subject;
(3) administration time for a single task should not exceed 30 minutes;
(4) each test should be sensitive for a large range of individual differences;
(5) each test test should measure a different skill;
(6) the order of administration should be unimportant;
(7) each test should result in one or two simple numbers as output;

(8) the scores should have a high reliability;
(9) learning effects should either be small or there should be reliable measures on samples of performance early in practice;
(10) each test should be well established with norms and reliabilities available in literature.

- reported validation procedures:
The authors recommend regression equations which can predict success in pilot training with multiple correlation coefficients ranging from 0.409 to 0.525.

- reported theoretical background for the whole battery:
The theoretical basis of this battery is information-processing theory.

(b) Specific description of subtests:

*** BBN 1: DIGIT SPAN TEST ***

- main references:
Stanford-Binet Test; WAIS

- theoretical background/ performance domain:
active memory capacity

- stimulus materials:
tape recorder, ear phones

- procedure:
The subject is asked to listen to a list of digits and to recall this list immediately in the correct order. This procedure is repeated with the number of digits per list being increased every second trial until the subject fails twice to reproduce lists of a given length correctly.
There are five sets of lists administered. The lists contain 4 to 12 digits. The stimuli are presented via tape recorder at a rate of 2 digits/sec.

- administration time:
15 minutes

- scoring and norms:
As the best measure of the subjects digit span the modal value of the (list lengths minus one) of the last correctly reproduced list in each set is calculated.

*** BBN 2: ROTATED LETTERS TASK ***

- main references:
Shepard & Metzler, 1971

- theoretical background/ performance domain:
spatial orientation; processing of spatial disparate sources of information

- stimulus materials:
paper and pencil, stopwatch

- procedure:
The subject's task is to distinguish between rotated letters and mirror images of the same letter.
Pairs of letters at 0, 50, 100, 150 degrees rotation disparity are presented to the subject, who has to decide whether both letters are the same (both normal or both mirrored).

- administration time:
15 minutes

- scoring and norms:
* the overall mean response time per item;
* slope of the best fitting regression line relating mean RT /item to angular disparity;
* percent correct responses.

*** BBN 3: DICHOTIC LISTENING TEST ***

- main references:
Gopher & Kahneman, 1971

- theoretical background/ performance domain:
selective attention

- stimulus materials:
tape recorder, ear phones

- procedure:
Subjects receive a sequence of dichotically presented digits and color names. They are asked to shadow those digits that occur on the so-called relevant ear. After 3 - 6 pairs of items the relevant ear is redefined. A high tone signals the right ear to be relevant, a low tone the left ear. The order of tones is randomly determined. The tones last 500 msecs with 500 msecs pause afterwards. The items are presented simultanuously at a rate of 2 items/sec.
A trial consisted of 4 blocks of item pairs. 3 training and 24 experimental trials are administered.

- administration time:
15 minutes

- scoring and norms:
* number of blocks where no (correct) response has been made (=missed block= 1 error) although there should have been at least one,
* omissions in blocks, where at least one correct response has been made
* digit intrusion (signals from irrelevant ear)
* color intrusion (signals from irrelevant ear)
* other mistakes,
* overall performance measure S = approximate percent correct:

S=(N-L-K)/(N-L)

N: total number of pairs
L: number of missed blocks times the average number of pairs per block = number of missed pairs
K: total number of errors of all types

*** BBN 4: STROOP TEST ***

- main references:
Stroop, 1939

- theoretical background/ performance domain:
perception, cognition

- procedure:
The stimulus set consists of 2 color cards (c-cards) and 2 color-word cards (cw-cards). On each card are 72 items in form of colored plastic stripes. On c-cards there are white Xs printed on the stripes, on cw-cards nonfitting colornames. The subjects are asked to name the colors of the stripes.

- administration time:
10 minutes

- scoring and norms:
* time needed to name colors on c-cards
* time needed to name colors on cw-cards
* the difference between these two scores

*** BBN 5: SENTENCE VERIFICATION TASK ***

- main references:
Chase & Clark, 1972; Trabasso, 1972; Wason, 1959

- theoretical background/ performance domain:
linguistic decoding

-procedure:
32 sentence/letters-pairs of the rootform "A precedes B/ AB" are presented to the subject who has to decide whether the sentence is a correct description of the subsequent letters.

- administration time:
5 minutes

- scoring and norms:
* time required to complete all 32 test items
* percent correct responses

*** BBN 6: CRITICAL TRACKING TASK ***

- main references:
Jex, McDonnell & Phatak, 1966

- theoretical background/ performance domain:
perceptual-motor performance

- stimulus materials:
special electronical tracking apparatus

- procedure:
The subject is asked to control a target spot on a visual display by means of a joystick. The target is programmed to move to the left or right unless a correcting control movement is introduced. During the trial the spot becomes more and more unstable, thus the subject has to react faster and faster to control it, until this becomes impossible.
Subjects are given three training trials followed by 7 practice trials.

- administration time:
15 minutes

-scoring and norms:
* mean value of the time constant tau of the system at the time of loss of control.

*** BBN 7: TIME SHARING - TRACKING AND DIGIT SPAN TEST ***

- main references:
see BBN 1 and BBN 6

- theoretical background/ performance domain:
time sharing

- stimulus materials:
analog to BBN 1 and BBN 6

- procedure:
The difficulty level of the tracking test is fixed at a moderate niveau (mean tau(crit) + 20 msecs). The length of the digit lists is constant, set equal to the individual digit span minus 1.
Each trial takes 65 secs: 5 secs tracking warming up, 30 secs tracking only, 30 secs simultanuously tracking and digit span test. 10 trials are performed. The subject is instructed to maintain a maximum level of performance in the digit span test and keep the target in the center of the display. The digits have to be recalled immediately after each trial.
If the subject loses the target off the border of the display the trial is terminated and rerun.

- administration time:
20 minutes

- scoring and norms:

For the tracking task the scores are measured in terms of the mean distance of the target from the center of the display ("integral-absolute-error"). Three scores are computed:

* the average for the 30 secs tracking only
* the average for time shared tracking time
* the difference between the two.

For the digit span test, the percentage of correct reported digits is obtained.

*** BBN 8: TIME SHARING - TRACKING AND DICHOTIC LISTENING ***

- main references:

see BBN 3 and BBN 6

- theoretical background/ performance domain:

time sharing

- stimulus materials:

tracking apparatus, tape recorder, ear phones

- procedure:

Subjects are instructed to maintain maximum performance on the dichotic listening test, and to keep the target in the center of the display. The procedure was analog to the last test: 5 secs tracking warming up, 30 sec tracking only, 30 sec time shared tracking and dichotic listening.

- administration time:

20 minutes

- scoring and norms:

Scoring is analog to the last test and the dichotic listening task:

* mean-integral-absolute-error for the tracking-alone intervals
* mean-integral-absolute-error for the timeshared-tracking interval
* difference betwen the two
* number of missed blocks
* omissions
* digit intrusions
* color intrusions
* other errors
* average percent correct.

5.3.3. CTS - CRITERION TASK SET

(a) General description of the battery

- authors:
Shingledecker, C.A.

- title:
A Task Battery for Applied Human Performance Assessment Research

- source:
Air Force Aerospace Medical Research Laboratory Report Number AFAMRL-TR-84-071

- reported original purpose:
"The theoretical basis and standardized features of the CTS make it potentially applicable to a number of research problems in the areas of human performance assessment and human factors. One of these problems for which the CTS was originally designed is the comparative evaluation of measures of mental workload. In this application, the individual components of the CTS are being used as primary loading tasks to assess the reliability, sensitivity and intrusiveness of a number of proposed behavioral, subjective, and physiological indices of workload. ... A second broad area of investigation to which the CTS can be applied as a standardized test instrument is the assessment of human performance capabilities. When used for this purpose, the tasks comprising the CTS may be employed in a diagnostic fashion to measure and predict the effects of extreme environments and biochemically active agents on human performance" (Shingledecker, 1984, 11f).

- reported criteria for the selection of subtests:
To guide the development of a set of tasks for the CTS the author summarized the state-of-the-art research findings and conceptual approaches in a theoretical model. Primary components of this model were derived from multiple resource theories and processing stage theories (e.g. Wickens, 1981; Sternberg, 1969).
Practical selection criteria were the ability to manipulate task demand levels and to minimize loading on resources not tested by the task and good face validity in order to enhance subject's acceptance of the task and to allow easiier generalization to real life tasks.

- reported theoretical background for whole battery:
mainly multiple resource and processing stage theories (e.g. Wickens, 1981; Sternberg, 1969).

(b) Specific description of the subtests:

All tasks are implemented in user-friendly software on an inexpensive microcomputer system with some additional custom-made hardware. The whole system consists of ten parts:
1. Commodore 64 microcomputer

2. Commodore 1541 diskdrive
3. Commodore C1526 printer
4. monochrome experimenter's monitor
6. experimenter's video monitor switch (custom)
7. Commodore 1702 color subject's monitor
8. four button response keypad (custom)
9. tapping key (custom)
10. rotary tracking control (custom)

Position 1 to 6 establish the experimenter's teststation, 7 to 10 the subject's.

*** CTS 1: PROBABILITY MONITORING ***

- main references:

Chiles, Alluisi & Adams, 1968

- theoretical background/ performance domain:

visual perception; scanning; detection; monitoring

- stimulus materials:

display, four button response keypad

- procedure:

The subject is asked to monitor one or more displays which have the appearance of electromechanical dials with pointers. Under the nonsignal condition the pointer moves randomly on the display. Under the signal condition it moves predominantly with a preselected probability (0.95, 0.85, 0.75) only on one side of the dial.
These biases in pointer movement are supposed to be signals for the subject to press an appropriate response key. The subjects are instructed not to respond until they are really sure there is a signal present. Each trial takes 3 minutes with 2 - 3 signals during that time. A minimum of 25 sec will separate two signals from eachother. Undetected signals will last 30 sec.
Task difficulty can be manipulated by varying the signal probability and the number of dials.

- administration time:

3 min/ trial

- scoring and norms:

* reaction time for correct responses
* number of false alarms
* number of missed responses

*** CTS 2: CONTINUOUS RECALL TASK ***

- main references:

Hunter, 1975

- theoretical background/ performance domain:
working memory encoding; memorizing; keeping track of events; recalling recent events

- stimulus materials:
display; four button responses keypad with only the two keys at the extreme left and right to be used.

- procedure
Simultaneously two random numbers are presented on a display: a test item and a probe item. The subject is instructed to encode the test item and to compare the probe item with a test item presented previously a number of positions back in the series and to decide whether it is the same by pressing an appropriate response key.
The task is subject paced with a preselected reaction time deadline. Task difficulty can be manipulated by varying item length and length of the item series which must be maintained in memory for the comparison of probe and test item. Three different demand levels are recommended. Subjects are instructed to react as fast and as accurately as possible.
Major practice effects can be eliminated with 3 - 7 training trials.

- administration time:
3 min/ trial

*** CTS 3: MEMORY SEARCH TASK ***

- main references:
Sternberg, 1969

- theoretical background/ performance domain:
working memory retrieval; memorizing; keeping track of events; recalling recent events

- stimulus materials:
display, four button response keypad with the extreme buttons to be used only

- procedure:
A small set of letters is visually presented to the subject for memorization (=memory set). Then a series of single letters is presented (test items). For each test item the subject has to decide whether it has been contained in the memory set.
The task is subject paced with a preselected reaction time deadline. Task difficulty can be manipulated by varying size of the memory set. Three levels are recommended (1, 2, 4 items/ set).
Major practice effects can be eliminated with 7 - 16 training trials.

- administration time:
3 min/ trial

*** CTS 4: LINGUISTIC PROCESSING TASK ***

- main references:
Posner, 1967

- theoretical background/ performance domain:
symbolic information manipulation; analysis of meaning; language comprehension; classification of events

- stimulus materials:
display, four button response keypad with the extreme buttons to be used only

- procedure:
The subject has to classify a pair of visually presented letters or words as matching/ notmatching on the basis of given classification rules by pressing an appropriate response key.
The task difficulty depends on the classification rule. Three levels are recommended:
   * physical identity classification
   * category match (both consonants or vowls)
   * antinym match.
Subjects are instructed to respond as quickly as possible without errors.
The task is subject paced with a deadline.
Major practice effects can be eliminated with 5 -10 practice trials.

- administration time:
3 min / trial

- scoring and norms:
percent errors

*** CTS 5: MATHEMATICAL PROCESSING TASK ***

- theoretical background/ performance domain:
symbolic information manipulation; computing; calculating; comparison of values

- stimulus materials:
display; four button response keypad with the extreme buttons to be used only

- procedure:
Subjects have to perform simple arithmetic operations on a number of visually presented digits and to decide whether the result is greater than a prespecified value by pressing an appropriate key. Subjects are instructed to operate from left to right. The task is subject paced with a deadline.
Task demands depend on the number and combination of operations required.
Three levels are recommended:
   * low: one operation / +, -
   * medium: two operations / +- or -+

* high: three operations / ++- or +-+ or -+-

- administrations time:
3 min / trial

- scoring and norms:
percent correct

*** CTS 6: SPATIAL PROCESSING TASK ***

- main references:
Chiles, Alluisi & Adams, 1968

- theoretical background/ performance domain:
spatial information manipulation; maintaining orientation; identifying patterns; analyzing positions

- stimulus materials:
display; four button response keypad with the extreme buttons to be used only

- procedure:
The subject has to view pairs of histograms sequentially presented and to decide whether they are identical by pressing an appropriate response key.
Task demand levels are manipulated by varying the number of bars and the spatial orientation of the second histogram.
Ten practice trials are recommended to eliminate practice effects.

- administration time:
3 min/ trial

- scoring and norms:
reaction time; percent correct

*** CTS 7: GRAMMATICAL REASONING TASK ***

- main references:
Baddeley, 1968

- theoretical background/ performance domain:
reasoning; problem solving; analyzing relationships; logical thinking

- stimulus materials:
display; four button response keypad with the extreme buttons to be used only

- procedure:
Stimulus items are one or more sentences accompanied by a string of symbols. The subject has to decide whether the sentences are a correct description of the symbol string.
Task demand depends on the number of sentences (symbols) and the syntactic

structure. Three levels are recommended. Nine training trials are recommended to eliminated practice effects.

*** CTS 8: UNSTABLE TRACKING TASK ***

- main references:
Jex, McDonnell & Phatak, 1966

- theoretical background/ performance domain:
manual response; speed accuracy; continuous control; error correction; control actuation

- stimulus materials:
display; rotary tracking control

- procedure:
Subjects are instructed to keep a vertically moving cursor in the center of a display by means of a joystick.
"The system represented by the task is an inherently unstable one. The operator's input introduces error which is magnified by the system with the result that it becomes increasingly nescessary to respond to the velocity of the cursor movement as well as to the cursor position". If the subject loses the target off the border of the display it returns automatically to the center.
Three demand levels are recommended. 1 - 12 training trials are recomended to eliminate practice effects.

- administration time:
3 min/ trial

- scoring and norms:
* average absolute tracking error
* number of controll losses

*** CTS 9: INTERVAL PRODUCTION TASK ***

- main references:
Michon, 1966

- theoretical background/ performance domain:
manual response timing; scheduling movements; coordinating squential responses

- stimulus materials:
tapping key

- procedure:
Subjects are instructed to do fingertapping at a rate of 1-3 taps per second. Four training trials are recommended.

- administration time:
3 min / trial

- scoring and norms:
* standard deviation of interval durations
* "IPT variability score"

5.3.4. IPT - Test battery

(a) General description of the battery:

- authors:
Rose, A.M.

- title:
Information processing abilities.

- source:
R.E.Snow, P.A. Frederico, & W.E. Montague (Eds.), Aptitude, learning and introduction. Hillsdale, NJ, 1980.

(For internal use this battery will be called IPT = information processing tasks).

- reported original purpose:
This test battery consists of a number of information processing tasks. It "is designed to be used as an assessment device for performance evaluation in the context of personnel management. Another application of this type of test battery includes assessing the effects of unusual environments on cognitive performance" (p.67).

- reported criteria for the selection of subtests:
The tasks were gleaned from the literature on information processing as representatives of well understood and empirically studied paradigms.
The tasks had to be adaptable to paper and pencil format or to small digital computers or to some other form that could easily be administered in a group setting.

- reported theoretical background:
information processing theory

(b) Specific description of subtests:

*** IPT 1: LETTER CLASSIFICATION TASK ***

- main references:
Posner & Mitchell, 1967

- theoretical background/ performance domain:
matching/ recognition at different levels of stimulus complexity.

- procedure:
Pairs of letters are presented to subjects who have to decide whether these letters are in a certain way the same or not.
For the first block of trials sameness is defined as physical identity (aa,AA,bb,...), for the second block as name-identity (aA,AA,Bb, ...), for the third block as category-identity, i.e.: both letters being vowls or both letters being consonants (AE,bc,DG,...).
These three classification rules seem to represent different task demand levels:
 * physical-identity rule: low
 * name-identity rule: medium
 * category-identity rule: high

- scoring and norms:
reaction time

*** IPT 2: LEXICAL DECISION MAKING TASK ***

- main references:
Meyer, Schvaneveldt & Rudy, 1974

- theoretical background/ performance domain:
recognition of written words.

- procedure:
On each trial two strings of letters are displayed successively and the subject has to decide whether they are English words or nonwords.
The critical variable is the graphemic and phonemic relation within each pair of words. There are three types of relations:
 * phonemically similar/ graphemically similar
 * phonemically similar/ graphemically dissimilar
 * phonemically dissimilar/ graphemically similar.

- scoring and norms:
reaction time for each string

*** IPT 3: GRAPHEMIC AND PHONEMIC ANALYSIS TASK ***

- main references:
Baron, 1973

- theoretical background/ performance domain:
differentiation between phonemic and graphemic encoders.

- procedure:
Subjects are asked to decide whether various presented sentences make sense. Three types of phrases are used:
   * sensephrases (S)
   * nonsense phrases (N)
   * phrases that sound sensible because of a homophone, but look like nonsense (H).

In a first block of trials S and H phrases are used and the subjects are instructed to classify the phrases on the basis of their appearance. In a second block H and N phrases are used and the classsification basis should be how they sound. In a third block S and N phrases are used and the subjects are allowed to use the basis they prefer.

- scoring and norms:
reaction times for each trial (rsp. blockwise by Baron).

*** IPT 4: SHORT TERM MEMORY SCANNING ***

- main references:
Sternberg, 1967, 1969

- theoretical background/ performance domain:
memory scanning

- procedure:
On each trial a list of randomly selected digits (1 - 9) is presented for memorization (memory set). After a short pause a single digit is presented (test stimulus) and the subject has to decide whether the test digit is a member of the memory set.

- scoring and norms:
reaction time from test stimulus onset to response.

*** IPT 5: MEMORY SCANNING FOR WORDS AND CATEGORIES ***

- main references:
Juola & Atkinson, 1971

- procedure:
This task is a sort of variation of the Sternberg paradigm, using sets of one to four words (esp. category labels) rather than digits.
Under the first condition a positive probe stimulus is one of the items (category names) from the memory set.
Under the second condition a positive probe stimulus is an instance from those categories.

- scoring and norms:
reaction time

*** IPT 6: LINGUISTIC VERIFICATION TASK ***

- main references:
Clark & Chase, 1972

- procedure:
On a display a picture and a sentence are shown. The subject has to decide whether the sentence is a true description of the picture.

- scoring and norms:
reaction time

*** IPT 7: SEMANTIC MEMORY RETRIEVAL ***

- main references:
Collins & Quillian, 1969

- theoretical background/ performance domain:
semantic memory retrieval

- stimulus materials:

- procedure:
To study the subject's access to hierachically organized information the subject has to decide whether a presented sentence--either of property- or subset-type--is true.

- scoring and norms:
reaction time

*** IPT 8: RECOGNITION MEMORY TASK ***

- main references:
Shephard & Teghtsoonian, 1961

- procedure:
Subjects are presented with a lengthy list of items. They are asked to identify each item as "new" or "previously presented". The interval between the original and the test presentation of the items is varied.

5.3.5. PAB - PERFORMANCE ASSESSMENT BATTERY

(a) General description of the battery

- title:
Performance Assessment battery (PAB)

- source:
Thorne, D., Genser, S., Sing, H., & Hegge, F.
Plumbing human performance limits during 72 hours of high task load.
DRG Seminar, Toronto 1983, Seminar Paper.

- reported original purpose:
The battery has been developed for military purposes at the Walter Reed Army Institute of Research WRAMC, Washington.

- reported criteria for the selection of subtests:
Some subtests were adapted from pre-existing paper and pencil tests, from memory-drum or tachistoscopic-type tests; others were developed specifically for this battery.

(b) Specific description of subtests:

*** PAB 1: TWO-LETTER-SEARCH ***

- main references:
Folkard et al., 1976

- theoretical background/ performance domain:
visual search; recognition

- stimulus materials:
computer display; keyboard

- procedure:
Two target letters are presented at the top of the screen, followed by a string of 20 letters in the middle of the screen. The subject determines as quickly as possible whether both target letters are present in the string or not. If both are present, in any order, the "S" key is pressed for "same". If one ore more letters are missing, the "D" key is pressed for "different".

- administration time:
2 minutes

*** PAB 2: SIX-LETTER SEARCH ***

- main references:
see above

- theoretical background/ performance domain:
visual search; recognition

- stimulus materials:
computer display; keyboard

- procedure:
Analog to the above described task, but with six target letters instead of two.
Evidence has been reported suggesting that the additional memory load associated with this task causes it to exhibit a different circadian pattern than the two-letter task.

- administration time:
2 minutes

*** PAB 3: TWO-COLUMN ADDITION ***

- stimulus materials:
computer display; keyboard

- procedure:
Five two-digit numbers are presented simultaneously in column format in the center of the screen. The subject calculates their sum as rapidly as possible and enters it from the keyboard, most-significant digit first. The column of digits disappears with the first key entry, and no aids for "carry" operations are allowed. The task is subject paced.

- administration time:
3 minutes

*** PAB 4: LOGICAL REASONING ***

- main references:
Baddeley, 1968

- theoretical background/ performance domain:
transformational grammar

- stimulus materials:
computer display; keyboard

- procedure:
The letter pair "AB" or "BA" is presented along with a statement that correctly or incorrectly describes the order of the letters within the pair (e.g.,"B follows A", or "A is not preceded by B"). The subject decides whether the statement is true (Same) or false (Different) and presses the "S" or "D" key accordingly. The "S" and "D" keys are chosen over the "T" and "F" keys because they are adjacent to one another on a conventional keyboard. The 32 possible sentence/ pair combinations are presented once each or until four minutes have elapsed.

- administration time:
no longer than 4 minutes

*** PAB 5: DIGIT RECALL ***

- theoretical background/ performance domain:
short term memory capacity

- stimulus materials:
computer display, keyboard

- procedure:
Nine random digits are displayed simultanuously in a row across the center of the screen for a second. After a three-second blank retention interval, eight of the original nine digits are re-displayed in a different random order, and the subject enters the missing digit. A given digit may appear no more than twice on each trial, although subjects are not informed or generally aware of this constraint.

- administration time:
3 minutes

*** PAB 6: SERIAL ADD/ SUBTRACT ***

- main references:
Pauli; Wever, 1979

- theoretical background/ performance domain:
sustained attention; machine-paced calculating

- stimulus materials:
display; kevboard

- procedure:
Two randomly selected digits and either a plus or minus sign are displayed sequentially in the same screen location followed by a prompt symbol. The subject performs the indicated addition or subtraction and enters the least significant digit of the result. If the result is negative he adds ten to it and enters the positive single digit remainder (e.g., 3 9 - equals -6, so enter 4). The digits and signs are presented for approximately 250 milliseconds, separated by approximately 200 milliseconds. The next trial begins immediately after the key entry.

- administration time:
no longer than 4 minutes or 50 trials

*** PAB 7: PATTERN RECOGNITION 1 ***

- theoretical background/ performance domain:
spatial memory

- stimulus materials:
computer display; keyboard

- procedure:
A random pattern of dots (asterisks) is displayed for 1.5 seconds and then followed after a 3.5 second retention interval by a second pattern that may be same or different. The subject has to press the "S"-key for same or the "D"-key for different. The pattern consists of 14 dots, of which either three or no dots change location.

- administration time:
Ten trials are run.

*** PAB 8: PATTERN RECOGNITION 2 ***

- theoretical background/ performance domain:
spatial memory

- stimulus materials:
computer display; keyboard

- procedure:
This task is a more difficult version of the above. The pattern consists of 16 dots, of which either two or no dots change. Ten trials are run.

*** PAB 9: LEXICAL DECISION TASK ***

- main references:
Babkoff; Genser & Babkoff.

- stimulus materials:
computer display; keypad; eye patch

- procedure:
The subject wears an eye patch and fixates the center of a CRT screen with head fixed in position by forehead and chin rests. Strings of three to five letters are displayed briefly on the screen either to the left or right visual field and the subject presses one of two buttons indicating whether the string was a word or a non-word.

- administration time:
20 minutes

*** PAB 10: VIGILANCE & DETECTION TASK ***

- main references:
Taube.

- stimulus materials:
video monitor; speech synthesizer; response key

- procedure:
A series of random digits randomly selected from "1" through "4" is rapidly presented either visually on a video monitor, vocally with a speech synthesizer, or both. The subject presses a button as quickly as possible every time the digit "3" occurs. The rate of stimulus presentation adjusts to the subject's reaction time and error rate.

- administration time:
5 minutes

*** PAB 11: ILLUSION SCALE ***

- theoretical background/ performance domain:
hallucinations etc.

- stimulus materials:
video monitor; ?

- procedure:
This task consists of the video presentation of 52 questions concerning sensory/ perceptual illusions, distortions and hallucinations, along with self assessments of motivation and performance which the subject scores on a five point scale.

- administration time:
3 minutes

*** PAB 12: FATIGUE CHECK LIST ***

- stimulus materials:
paper and pencil

- procedure:
This task consists of 30 forced choice questions dealing mostly with possible somatic complaints.

- administration time:
1 minute

*** PAB 13: MOOD ACTIVATION SCALE ***

- main references:
Genser; Thayer, 1967; Zuckerman, 1964 & 1965.

- theoretical background/ performance domain:

- stimulus materials:
either video monitor and keyboard or paper and tape recorder

- procedure:
Subjects are presented with 65 adjectives and are asked to respond on a

five point scale with the extent to which the adjectives reflect their current feelings. The adjectives were selected to represent positive affect, or feeling "good"; negative affect or feeling "bad"; positive activation, or feeling "energetic"; and negative activation, or feeling "tired". Examples of each category are "happy, cheerful/ sad, mad/ active, alert/ sleepy, drowsy", respectively.
The adjectives were either presented one-by-one on a video monitor and responded to manually by keyboard, or they were presented as a list on a printed page and responded to orally by dictating into a tape recorder.

- administration time:
3 minutes

5.3.6. TTP - TEN TASKS PLAN - TASKOMAT

(a) General description of the battery

- authors:
Boer, L.C. & Gaillard, A.W.K.
Institute for Perception, TNO, Soesterberg, NL

- title:
TASKOMAT - A Standardized Task Battery

- source:
unpublished paper, April 1986, and personal report of the authors.

- reported original purpose:
The task battery may be used for (1) the selection of personnel, (2) the evaluation of training, (3) the assessment of stressor effects, (4) the measurement of mental fitness, (5) as an estimate for mental workload. Either intra- or interindividual differences can be assessed.

- reported criteria for the selection of subtests:
(1) currency in the human-performance literature;
(2) specific measures with a high contruct validity;
(3) have been applied several times by the TNO Institute for Perception--this implies validation sudies, showing effects of fatigue, stress, psychoactive drugs, and criterion-related individual differences in skills;
(4) feasibility for a task battery (ad-ministration time, technical requirements).

- reported validation procedures:
All tasks, except 3 and 4, have shown some validity in the past. Tasks 1 and 2 have shown effects of sleep deprivation and fatigue (Boer et al., 1984; Sanders et al., 1982), and task 1 has shown selective effects of drugs (Frowein, 1979, 1981; Frowein et al., 1981; Gaillard & Verduin, 1983) and of brain damage (Stokx & Gaillard, 1986). Tasks 5-8 have shown predictive validity for flight training (Gaillard et al., 1984; Gopher,

1982). Additionally task 5 has shown effects of stress and workload of divers, and the reduction of these effects after training (Jorna, 1981, 1982). The task also discriminated between aviator groups differing in proficiency (Boer, 1986).
Validation of the tasks in the present task-battery form is in progress. Drug effects on tasks 1 and 6 have been assessed. No effects were observed for task 1. For task 6 effects were reliable (Gaillard, 1986).

- reported theoretical background for whole battery:
The background research of the TTP relies strongly, although not exclusively, on processing stage descriptions of choice reactions.

(b) **Specific description of subtests:**

The visual tasks are implemented on an IBM PC/XT.

*** TTP 1: **RT task** ***

- theoretical background/ performance domain:
additive factor theory (Sternberg, 1969; Sanders, 1980)

- stimulus materials:
computer display; four button response keypad

- procedure:
A stimulus (digits 2,3,4,5) is shown either on the left or right of the screen. The subject has to press a corresponding response key with his index or middle finger of either his left or right hand.
There are four task variables which are varied separately in "blocks":
(1) stimuli intact vs. stimuli degraded, (2) S-R compatibility/ incompatibility, (3) single response/ complex response (sequence of three keys to be pressed instead of one), (4) fixed interstimulus intervals of 2 secs vs. "time uncertainty" with ISIs of 2 - 10 secs.

- administration time:
blocks å 4 min.

- scoring and norms:
RT
--Normative data are currently being collected.

*** TTP2: MEMORY SEARCH TASK ***

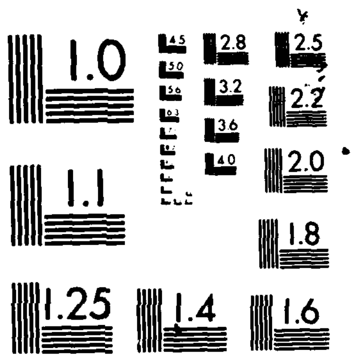
- theoretical background/ performance domain:
memory search (Sternberg, 1969a); automatic versus controlled processing (Shiffrin & Schneider, 1977)

- stimulus materials:
computer display; response key

AD-A172 682 STANDARDIZATION OF PERFORMANCE TESTS: A PROPOSAL FOR FURTHER STEPS(U) TECHNISCHE HOCHSCHULE AACHEN (GERMANY F R) INST FOR PSYCHOLOG A F SANDERS ET AL 01 JUL 86 2/2

UNCLASSIFIED EOARD-TR-86-08 AFOSR-85-0305 F/G 5/10 NL

END



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

- procedure:
In this task subjects have to decide whether a display contains a "target". One to four symbols (digits, upper-case letters) are simultanuously presented. They are positioned in a small 2X2 matrix. Vacant positions, if any, are filled with plus signs.
The task is self paced. Stimuli are presented in blocks. Each block starts with a message on the screen telling the subject which symbols are targets (1-4). For yes-responses the subject has to press a button with his right hand, for no-responses another one with his left hand.
The task variables are: (1) number of stimulus elements in the matrix, (2) number of targets, (3) whether or not there is a categorical distinction between targets and other elements.

- administration time:
blocks à 4 minutes

- scoring and norms:
reaction time--both for targets and non-targets--is the major dependent variable. It is calculated as a function of the number of stimulus elements (display load).
--Normative date are currently being collected.

*** TTP 3: SELECTIVE-ATTENTION-TASK ***

- theoretical background/ performance domain:
automatic versus controlled processing (Shiffrin & Schneider, 1977); focussing attention (Eriksen & Schultz, 1979); Okita et al. 1985

- stimulus materials:
computer display; response keypad

- procedure:
In this task 1-4 symbols (digits, upper-case letters) are presented in a small 2X2 matrix. Vacant positions, if any, are filled with plus signs. One diagonal of the matrix is defined as relevant, the other as irrelevant. The subject is instructed to attend selectively to the relevant diagonal and to detect eventual targets on this diagonal.
The task is self paced. Stimuli are presented in blocks. Each block starts with a definition of the targets. "Yes"-responses are performed with the right hand, "no"-responses with the left hand. Task variables are: (1) distraction value of the unattended diagonal, i.e. whether there are plusses, letters, or "targets" on this diagonal, (2) number of targets, (3) whether there is a categorical distinction between targets and nontargets.

- administration time:
blocks à 7 minutes

- scoring and norms:
reaction times for yes- and no-reactions, as a function of the distraction value of the unattended diagonal.
--Normative data are currently being collected.

*** TTP4: RESPONSE CONFLICT TASK ***

- theoretical background/ performance domain:
focussing attention (Eriksen & Schultz, 1979); positional compatibility (Simon et al.,1976)

- stimulus materials:
computer display; response keypad

- procedure:
Stimulus elements are the upper-case letters A and B. Prior to stimulus presentation a 500 ms fixation mask is presented, which marks the positions of the stimulus elements. The subject has to press the left key if an A is presented in a critical position, the right key if a B is in the critical position. Presentation of stimuli is self paced.
In the "position certain" block, three letters are presented. Critical is the letter in the middle of the triple. The two flanking letters are to be ignored. Sample stimuli include AAA, BAB.
In the "position uncertain" block the stimulus consists only of one letter, A or B. The letter may either be in the left or right position, and may be flanked by a digit from the set 3,4,6,7,9. The subject has to press the key corresponding to the letter.
Task variables are (1) positional certainty/ uncertainty, (2) for "position certain" blocks the amount of conflict between critical letter and flanking letters, and, (3) for "position uncertain" blocks positional compatibility between the critical letter and the correct response key.

- administration time:
blocks à 2.5 minutes

- scoring and norms:
RTs

*** TTP 5: CONTINUOUS MEMORY TASK ***

- main references:
Massaro, 1975; Sternberg, 1969a; Shiffrin & Schneider, 1977.

- theoretical background/ performance domain:
mental workload; memory search; controlled processing

- stimulus materials:
taperecorder with computer-synthesized letters of the alphabeth; handheld response key

- procedure:
With mean interstimulus intervals (ISIs) of 2.25 secs (range: 1.5-4.5 secs) a sequence of consonants is auditorily presented. The subject's task is to indicate the occurrence of predefined targets by pressing a key and to count silently the number of occurrences separately for each target. The task variable is the number of targets, which is either two or four. One quarter of the stimuli are targets. At the end of each block the subject is

asked to report the sum of occurrences for each type of target.

- administration time:
blocks à 5 min.

- scoring and norms:
* deviation between actual and reported frequency of targets
* RT
* counting errors

*** TTP 6: TRACKING ***

- theoretical background/ performance domain:
tracking, anticipation of future control, skill development (Hess, 1981)

- procedure:
The tracking task consists of pursuit tracking of a sawtooth track within the boundaries of a window. To give the subject preview a part of the upcoming track is displayed in advance.
The cursor is a small horizontal line with a gap in the middle. The subject's task is to move the cursor horizontally by means of a control stick in such a way that the track passes through the middle of the gap without touching. Task variables are (1) the amount of preview, (2) speed of the track.

- administration time:
7 min blocks

- scoring and norms:
* root mean squared error
* number of times out of cursor line

*** TTP 7: DUAL TASK ***

- theoretical background/ performance domain:
dual task performance resource theory; POC function (Wickens, 1980; Norman & Bobrow, 1975)

- procedure:
This is a combination of the tasks TTP5, continuous memory, and TTP6, tracking. It is presented in triple blocks: tracking only, dual task, tracking only. In the dual task condition the instruction emphasizes the tracking task; tracking starts and the memory task is switched on one minute later.

- administration time:
three blocks of 21 minutes

- scoring and norms:
* root mean squared error
* number of times out of cursor line

—Normative data have been collected.

*** TTP 8: DICHOTIC LISTENING TASK ***

- main references:
Gopher & Kahneman, 1971

- theoretical background/ performance domain:
focussing and switching attention
- stimulus materials:
taperecorder with computer-synthesized one-syllable letters and digits

- procedure:
Simultaneously two different messages (sequences of consonants mixed up with a few digits) are played to each ear. A preceding signal tone indicates which ear is to be attended (high tone = right ear, low tone = left ear). A trial consists of a first indicator tone, 16 pairs of dichotic stimuli, a second indicator tone and 3 - 5 final dichotic pairs. The stimulus pairs are presented every 500 ms. The subject's task is to write down as many digits of the attended message as possible.

- administration time:
approx. 20 minutes

- scoring and norms:
intrusion and omission errors
—Normative data are available.

5.3.7. HAK - HÄKKINEN TEST BATTERY

Short description of the battery

- authors:
Häkkinen, S.

- source:
Trafic accidents and professional driver characteristics: a follow-up study, Acid. Anal. & Prev. Vol.11 pp. 7- 18

- reported original purpose:
This test battery has been developed to study the role of personal factors in trafic safety.

List of subtests:

H1. Square Test
H2. Path Training Test
H3. Mechanical Comprehension Test
   - these three tests are paper and pencil tests with emphasis on reasoning and space perception.

H4. Tapping
H5. Fork
   - these tests concern simple motor speed, reaction time and two hand coordination.

H6. Clock Test
   - tests attention span, anticipation, correct timing.
H7. Driving Apparatus Test
   - the subject has to keep a stylus, which is moved by a steering wheel , on a "highway", while he/she simultanuously has to react with hand/ foot movements to 4 different kinds of stimuli. Driving experience is irrelevant for this test.
H8. Expectancy Reaction Test
   - this test is a visual disjunctive reaction test. The subject has to react to certain stimuli with a simple hand movement. - to make it more difficult, there are distractors internal and external to the test.
   The test is designed to study whether the motor performance of a subject is relatively higher than his speed of perception.

Six personality test complete the battery.

REFERENCES

Baddeley, A.D. (1968). A 3-minute reasoning test based on grammatical transformation. Psychonomic Science, 10, 341-347.

Baddeley, A.D. (1979). Die Psychologie des Gedächtnisses. Stuttgart: Klett-Cotta.

Baron, J. (1973). Phonemic stage not necessary for reading. Quarterly Journal of Experimental Psychology, 25, 241-246.

Boer, L.C. (1986). Attention tasks and their relation to aging and flight experience. Soesterberg, NL: Institute for Perception TNO (IZF 1986-4).

Boer, L.C., Ruzius, M.H.B., Mimpen, A.M., Bles, W., & Janssen, W.H. (1984). Psychological fitness during manoeuvre. Soesterberg, NL: Institute for Perception TNO (IZF 1984-17).

Clark, H.H., & Chase, W.G. (1972). On the process of comparing sentences against pictures. Cognitive Psychology, 3, 472-517.

Chase, W.G., Clark, H.H. (1972). Mental operations in the comparison of sentences and pictures. In: L.W. Gregg (Ed.), Cognition in Learning and Memory.New York: Wiley.

Chiles, W.D., Alluisi, E.A., & Adams, O.S. (1968). Work schedules and performance during confinement. Human Factors, 10, 143-196.

Collins, A.M., & Quilillian, M.R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Cooper, L.A., & Shephard, R.N. (1973). Chronometric studies of the rotation of mental images. In: W.G. Chase (Ed.), Visual information processing. New York.

Edwards, W. (1966). Optimal strategies for seaking information: Models for statistics, choice reaction and human information processing. Journal of Mathematical Psychology, 2, 312-319.

Eriksen, C.W., & Schultz, D.W. (1979). Information processing in visual search: A continuous flow model and experimental results. Perception and Psychophysics, 25, 249-263.

Fitts, P.M., & Peterson, J.R. (1964). Information capacity of discrete motor responses. Journal of Experimental Psychology, 67, 103-112.

Folkard, S., Knauth, P., Monk, T.H., & Rutenfranz, J. (1976). The effect of memory load on the circadian variation in performance efficiency under a rapidly rotating shift system. Ergonomics, 19.

Frowein, H.W. (1979). Effects of amphetamine on response in choice reaction tasks. Soesterberg, NL: Institute for Perception TNO (IZF 1979-8).

Frowein, H.W. (1981). Selective drug effects on information processing. Dissertation, Katholieke Hogeschool Tilburg, NL.

Frowein, H.W., Gaillard, A.W.K., & Varey, C.A. (1981). EP components, visual processing stages, and the effect of a barbiturate. Biological Psychology, 13, 239-249.

Gaillard, A.W.K., Boer, L.C., & Jorna, P.G.A.M. (1984). The application of laboratory tasks in military settings. In: NATO Defence Research Group Panel VIII, Workshop "Applications of systems ergonomics to weapon system development", Vol.1, C77-C83. Shrivenham, U.K.

Gaillard, A.W.K., Gruisen, A., & de Jong, R. (1986). The influence of Loratadine (Sch 29851) on human performance. Soesterberg, NL: Institute for Perception TNO.

Gaillard, A.W.K., Houtmans, M.J.M., & Rozendaal, A.H. (1984). De taakprestatie van revalidanten met diffuse hersenbeschadiging. Soesterberg, NL: Institute for Perception TNO (IZF 1984-14).

Gaillard, A.W.K., & Verduin, C.J. (1983). The combined effects of an antihistamine and pseudo-ephedrine on human performance. Journal of Drug Research, 8, 1929-1936.

Gopher, D. (1982). A selective attention test as a predictor of success in flight training. Human Factors, 24, 173-183.

Gopher, D., Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. Perceptual and Motor Skills, 33, 1335-1342.

Imhoff, D.L., & Levine, J.M. (1981). Perceptual motor and cognitive performance task battery for pilot selection (AFHRL-TR-80-27). Brooks Air Force Base, Texas: Manpower and Personnel Division, Air Force System Command.

Hunter, D.R. (1975). Development of an enlisted psychomotor/perceptual test battery. AFHRL-TR-75-60, Air Force Human Resources Laboratory, Personnel Research Division.

Jex, H.R., McDonnell, J.D., & Phatak, A.V. (1966). A "critical" tracking task for manual control research. IEEE Transactions on Human Factors in Electronics, HFE-7, 138-145.

Jorna, P.G.A.M. (1981). Stress, information processing and diving. Soesterberg, NL: Institute for Perception TNO (IZF 1981-4).

Jorna, P.G.A.M. (1982). Stress, information processing and coping with depth: problems with individual divers. Soesterberg, NL: Institute for Perception TNO (IZF 1982-14).

Juola, F.J., & Atkinson, R.C. (1971). Memory scanning for words versus categories. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 522-527.

Kaufmann, E.L., Lord, M.W., Reese, T.W., & Volkmann, J. (1949). The discrimination of visual number. American Journal of Psychology, 62, 498-525.

Massaro, D.W. (1975). Experimental psychology and information processing. Chicago: McNally.

Meyer, D.E., Schvaneveldt, R.W., & Ruddy, M.G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. Memory & Cognition, 2, 309-321.

Michon, J.A. (1966). Tapping regularity as a measure of perceptual motor load. Ergonomics, 9, 401-412.

Navon, D. (1977). Forest before trees: the presence of global features in visual perception. Cognitive Psychology, 9, 353-383.

Norman, D., & Bobrow, D. (1975). On data limited and resource limited processing. Journal of Cognitive Psychology, 7, 44-60.

Okita, T., Wijers, A.A., Mulder, G., & Mulder, L.J.M. (1985). Memory search and visual spatial attention: An event-related brain potential analysis. Acta Psychologica, 60, 263-292.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). The measurement of meaning. Urbana, Ill.: University of Illinois Press.

Pew, R.W., Rollins, A.M., Adams, M.J., & Gray, T.H. (1977). Development of a testbattery for selection of subjects for ASPT- experiments. Bolt, Beranek & Newman, Inc. Report No. 3585.

Posner, M.J. (1978). Chronometric explorations of mind. Hillsdale, NJ: Erlbaum.

Posner, M.I., & Mitchell, R.F. (1967). Chronometric analysis of classification. Psychological Review, 74, 392-409.

Poulton, E.C. (1974). Tracking skills and manual control. New York:

Academic Press.
Poulton, E.C. (1981). Human manual control. In: V.B. Brooks, (Ed.), Handbook of physiology: The nervous system. Vol.2. Betheda: American Psychological Society.
Rose, A.M. (1980). Information-processing abilities. In: R.E. Snow, P.-A.Frederico & W.E. Montague (Eds.), Aptitude, learning and instruction. Hillsdale, NJ.
Sanders, A.F. (1980). Stage analysis of reaction processes. In: G.E. Stelmach, & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.
Sanders, A.F., & Borselen, J.W.van (1966). Continuing meomory and information processing. Journal of Experimental Psychology, 71, 844-848.
Sanders, A.F., Wijnen, J.L.C., and van Arkel, A.E. (1982). An additive factor analysis of the effects of sleep-loss on reaction processes. Acta Psychologica, 51, 41-59.
Shephard, R.N., Metzler, J. (1971). Mental rotation of three dimensional objects. Science, 171, 701-703.
Shepard, R.N., & Tegtsoonian, M. (1961). Retention of information under conditions approaching a steady state. Journal of Experimental Psychology, 62, 302-309.
Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.
Shingledecker, C.A. (1984). A task battery of applied human performace assessment. Air Force Aerospace Medical Research Laboratory, Report No.AFAMRL-TR-84-071.
Simon, J.R., Acosta, E., Mewaldt, S.P., & Speidel, C.R. (1976). The effect of an irrelevant directional cue on choice reaction time: The duration of the phenomenon and its relation to stages of processing. Perception and Psychophysics, 19, 16-22.
Sternberg, S. (1967). Two operations in character recognition: some evidence from reaction-time measurements. Perception & Psychophysics, 2, 45-53.
Sternberg, S. (1969a). Memory scanning: Mental processes revealed by reaction time experiments. American Scientist, 57, 421-457.
Sternberg, S. (1969b). The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and Performance II. Amsterdam: North Holland.
Sternberg, S. (1975). Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 27, 1-32
Sternberg, S., Kroll, R.L., & Wright, C.E. (1978). Experiments on temporal aspects of keyboard entry. In: J.P. Duncanson, (Ed.), Getting it together: Research and application in human factors. Santa Monica, CA: Human Factors.
Stevens, S.S. (1975). Psychophysics. New York: Wiley.
Stokx, L.C., & Gaillard, A.W.K. (1986). Task and driving performance of patients with severe concussion of the head. Journal of Clinical Experimental Psychology.
Stroop, J.R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18,643-662.
Thayer, R.E. (1967). Measurement of activation through self-report. Psychological Reports, 20.
Wever, R.A. (1979). The circadian system of man. New York: Springer.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (ed.), Attention and Performance VIII. Hillsdale, NJ: Erlbaum.

Wickens, C.D. (1981). Processing resources in attention, dual task performance, and workload assessment. Technical Report EPL-81-3, Engeneering Psychology Research Laboratory, University of Illinois at Urbana-Champaign.

Wickens, C.D. (1984). Engeneering psychology and human performance. Columbus, OH: Merrill.

Yntema, D.B., & Shulman, G.M. (1967). Response selection in keeping track of several things at once. In: A.F. Sanders (ed.), Attention and performance. Amsterdam.

Zukerman, M., Lubin, B., Vogel, L., & Valerius, E. (1964). Measurement of experimentally induced affects. Journal of Consulting Psychology, 28.

Zukerman, M., Lubin, B., Robins, S. (1965). Validation of the Multiple Affect Adjective Check List in clinical situations. Journal of Consulting Psychology, 29.

5.4. LITERATURE REVIEWS

5.4.1. MANUAL TRACKING
by Will Spijkers

THE DOMAIN OF TRACKING

The main characteristic of tracking tasks is that they require continuous control of some input (McCormick and Sanders, 1982). They belong to the domain of continuous manual control tasks in which an analog time-space trajectory is a critical feature. Human performance in manual control has been considered from two quite different perspectives: the skill approach and the tracking approach. The skill approach has primarily considered analog motor behavior in circumstances with little environmental uncertainty and relatively little training in situations where more or less the same response is required from trial to trial. An example is the execution of an aiming movement towards a particular target in response to a discrete signal. In contrast, the tracking approach examines human abilities in controlling dynamic systems to make them conform with certain time-space trajectories in the light of environmental uncertainty (Kelly, 1968; Poulton, 1974). For example, keeping your car in the right lane of a winding road.

TRACKING: MAJOR CONCEPTS AND DEFINITIONS

In tracking it is the task of the operator to make a system respond in correspondence to a desired goal. In present day laboratory, a tracking task is typically implemented on a computer in which the subject or human operator (HO) controls a system whose dynamics are computer simulated, by manipulating a control stick and observing the response as a moving symbol on a visual display.

Besides various limitations of the human operator, four task elements can be distinguished in a tracking task which influence the performance of the HO. These are (1) the input or desired trajectory of the system, (2) the display or the means whereby the operator views or hears information concerning the desired and actual state of the system, (3) the control device whereby the HO provides the system with input, (4) the dynamics of the system itself. The tracking loop is defined by the following four elements: Display, human operator, control-device and system or processes.

Before discussing in greater detail the contribution of these different elements of a tracking task to human tracking performance and the "tools" which have been developed in order to make the task easier for the HO and to improve his performance, the elements of the tracking loop will be briefly described in order to provide a frame of reference for that discussion.

Display. Depending on the display, the input signal (commanded input or target) can be viewed together with the system's output (controlled element/variable, cursor, follower). When both the actual target and cursor are displayed, it is called a pursuit display. If only the difference between the trajectories of target and cursor are shown the display is called compensatory. With a compensatory display the HO only knows how far the state of the system is from the desired state, he doesn't know the actual values of either cursor or target.

Human Operator. The human operator must be able to perceive the output of the system and decide on the basis of this output and of his knowledge of the system whether a corrective response has to be generated or not. The eventual response generated depends not only upon his knowledge of the system's state, but also on the perceptual, decisional and motor qualities of the HO. It is beyond the scope of this section to treat in extenso the capabilities and limitations of human information processing. Those will be dealt with only in so far it is of concern for tracking performance. For a more detailed treatment of human information processing the reader is referred to other contributions of the report.

Control device. When it is decided to respond a control has to be operated in order to provide the system with a certain input change. A control is any device that allows a human to transmit information to a machine. Three basic classifications of controls can be discerned regardless of the physical implementation of the control device: (1) discrete versus continuous operation, (2) linear versus rotary operation and (3) one- versus two-dimensional operation. Physical quality of the control affects the ease at which a control can be operated (i.c. required force, discriminability from other controls, shape, size). Far more important for human tracking performance are the relation between the action to be executed and the effect produced in the control, referred to as control dynamics and the correspondence between signal change and response to be made referred to as Stimulus-Response Compatibility.

Process or System. The output of a control device is fed into the machine. The dynamics of the system itself determine the output it will generate and feed back to the HO by means of a display. The mathematical relation between the input and the output of a system is described in a transfer function. Models of tracking behavior have used transfer functions to describe human performance. It appears that the limits of human tracking performance depend in important ways upon the transfer function of the system being controlled.

In the next sections, several aspects of tracking will be further pursued. A more detailed description will be given of contributions on tracking performance of the above mentioned components of man-machine systems such as : Types of input, system-order, displays, and controls. This is followed by a section on the limits of the HO. Next something will be said about models of the HO and multiaxis control. Before arriving at the conclusion section, attention is paid to some dependent measures which are commonly be used in the evaluation of tracking performance.

COMPONENTS OF THE TRACKING TASK AND THEIR EFFECTS ON PERFORMANCE

TYPES OF INPUT

Four types of inputs are commonly distinguished: step, ramp, sine and complex inputs.

Step input. When the track suddenly jumps to a new position, one has a step input. Examples are changing lanes while driving along a multilane highway, aiming a camera at an object, reaching to and placing a finger on a push button.

Large movements take relatively longer to complete than small movements, despite the faster rate of long movements. With respect to movement speed and accuracy, Fitts' law states that for medium range movements the movement time depends upon the distance to be travelled (A) and the accuracy (W=width of the target): MT = log2 (2A/W). Accuracy of a movement is also affected by the availability of visual guidance (Keele and Posner, 1968; but see Schmidt, 1982 for the effects of strategy). Based on the results of Keele and Posner it is generally accepted that up to movement durations of about .25 sec, accuracy of movements is little affected when visual feedback is omitted.

Reaction time for a correction varies from .5 sec down to almost no time at all. Long reaction times are found when subjects are not expecting the correction (see also Poulton, 1980).

Ramp inputs. In contrast to step inputs, the tracks in ramp inputs only gradually arrive at their new position; there is a continuous change during a certain period. Following a horserace with a tv camera from the center court or bringing a ship from its present route to the one parallel to it are illustrations of ramp inputs.

Tracking a constant rate ramp with a lever position control, shows an average correction rate of two per sec. This error correction time of .5 sec comprises a RT of about .25 sec followed by a movement taking the remaining .25 sec. Doubling the track rate does only slightly affect the response frequency so that the mean error is about twice as large.

The theoretical importance of ramp tracks with constant rates is that they set the operator a constant unchanging problem. Changes in performance must therefore be determined by his limitations, and by the strategies which he uses to overcome his limitations. Experiments with ramp tracks can indicate the nature of the operator's limitations and strategies.

Reaction times of the first response to velocity ramps are longer for slow than for fast ramps. Presumably it takes the operator more time to see a change in position when the velocity is smaller. Lag and lead errors are mostly seen early in training, although with third order movements lag errors remain present. Accuracy of rate tracking (first-order control) is also affected by the angle of movement direction. The notion control order will be explained later on.

With regard to multiple ramp tracks, for example a saw-tooth pattern, two kinds of range effects occur: 1) the overshoots at reversals are less when the reversals are located towards the top or bottom of the display, 2) repetitions of the same ramp show less overshoot than alternations since subjects learn to expect ramps of about the same length.

Sine wave and complex tracks. Tracking a sine wave implies adaptation to an ever changing velocity pattern. A track consisting of several sines of different frequency and amplitude can be considered as a complex input signal. Doubling the amplitude of a sine wave track about doubles the average error. Tracks with top frequencies of .2 Hz and below are easy to follow with a position control. Up to 1 Hz the performance does not deteriorate much. As the top frequency of a quasirandom track is increased, the average amplitude of the response decreases, the average time lag increases and the average amplitude of the remnant increases. The notion of remnant will be discussed later; at this point it suffices to say that a low remnant indicates succesful predicting of the track and preprogramming of responses.

For a more detailed discussion of step, ramp and sine wave inputs the reader is referred to Poulton (1974; 1980).

PROCESS OR SYSTEM

When the position of a control device is changed the system will bring about a change in output. The relation between the input given and the output produced is determined by the control dynamics and involves three features: time, ratio and order.

Time. If a certain time elapses between the moment that the input command is issued and the output appears, the system is said to introduce a delay or transmission lag. There exist various types of delays introduced by the system under control which belong to the category control system lags. A pure transmission lag delays the input but reproduces it in identical form T seconds later. Pure time delays are universally harmful in tracking, and tracking performance gets progressively worse with greater delays. The reason is apparent: corrective input must be based upon the future rather than on the present value of the error.

An exponential lag is characterized by a gradual arrival of the system's output on a commanded input. Normally an exponential lag is defined by its time constant T(i) which is the time that the output takes to reach 63% of its final value. Effects of exponential lags are often less harmful than those of pure time delays, because it is in a sense a combination of a zero-order and a first-order control. Immediately following step input, the response of the exponential lag looks very much like that of the velocity control system, later that of time delay. Furthermore, when controlled with high frequency corrections, a system behaves like a first-order system and these systems have substantial advantages over systems of zero-order. It are these advantages that prevent exponential lags controlled at higher frequencies from exerting the kinds of harmful effects that the pure time delay does.

A response property which is typical of many dynamic physical systems with mass and spring constants, is the second-order lag. In this case the system's output reaches the commanded (step) input after a considerable oscillation.

Effects of various types of lags are intrically related to the various features of the tracking system. Poulton (1974; p 373) points out that all three types of control lag increase error in tracking. But, it can have also beneficial effects. Poulton (1974; p 378) indicates that a design engineer may be able to reduce the effective order of a system from second-order or acceleration control to a first-order or rate control by introducing an approximately exponential lag.

A display time lag consists of a delay in both the track input and the output of the system in question. A delay not caused by the system is the response lag or the operator's effective time delay. This is the time taken by the operator to make a response to an input.

Ratio. A pure-gain element describes the ratio of the amplitude of the output to that of the input. High-gain systems are highly responsive to a minor change in control device. An example of a relatively low-gain system is one often applied for tuning an audio receiver: it takes several full turns of the knob to travel from one end of the scale to the other. This example is often used to illustrate the concept of Control-display ratio (C/D ratio) which descibes the relation between the amount of movement of a control and the resulting movement of the cursor on the display. It is preferable to describe this here than under the heading of displays as is generally done, because the system is responsible for this gain-factor. The advantage of a high C/D ratio is that the so called travel time is short because reaching the desired position requires only a small movement of the control. It will be evident that through the sensitivity of the cursor to a minimal movement of the control, a high C/D ratio is less desirable when a fine adjustment of the cursor is desired. Thus, a high C/D ratio results in a long adjustment time and so eliminates the benefits of a short travel time. It should therefore be the aim to select a C/D ratio which minimizes the sum of these two times.

Control order. A change of input must be tracked by means of a change in a control device. The effect of this change depends on the hierarchical organisation or control order of the system. For example, if it is intended to change lanes of a car a steering wheel movement is required. If the wheel is turned to the left, the direction of the car is leftward and continues to be so as long as this wheel position is held. The action of changing lanes requires twice a temporary deflection of the steering wheel followed by a return to the old starting position: one to leave the present lane and one to bring the car back into the straight ahead position again. The amount of deflection of the steering wheel, i.e. the amplitude of the movement, determines the velocity at which lateral position is changed. Because three movements are involved the steering action is called a second-order or acceleration control.

Control order refers to what we might call the hierarchy of control relationships between the movement of a control and the output it is intended to control (McCormick and Sanders, 1982). The more levels or control loops which are serially involved to bring a change about in the environment, the higher the order of the system. Various types of control can be distinguished, but for simplicity we will restrict ourselves to the more important ones: Position-, rate- and acceleration- control and differentiator. An extensive discussion of the various kinds of control orders can be found in Frost (1972).

Position (zero-order) control. In a position-control tracking task the movement of the control device controls its output directly, such as moving a spotlight to keep it on the actor on a stage or following a moving curved line with a pen or other device. If the system involves a display, there is a direct relationship between the control movement and the display movement it produces.

Rate (velocity or first-order) control. With a rate-control system the direct effect of the operator's movement is to control the rate at which the output is being changed. The lateral position of a wheelbarrow, is an example of a first-order system because the amount of force applied determines the rate at which the lateral position is changed. One needs just one movement in order to accomplish a change in lateral position.

Acceleration (second-order) control. Acceleration is the rate at which there is change in the rate of movement of something. Operation of the steering wheel of an automobile is an example of acceleration control since the angle at which the wheel is turned controls the angle of the front wheels. In turn, the direction in which the automobile wheels point determines the rate at which the automobile turns. Thus, a given rotation of the steering wheel gives the automobile a corresponding acceleration toward its turning direction.

Differentiator. In isolation, differential control systems are not frequently observed. However, they are of critical importance when they are placed in series with systems of higher order. They can reduce the effective order of the system by "canceling" one of the integrators and so make it easier to control.

The effect of system order on all aspects of performance may be best described in the following terms: zero-order and first-order systems are roughly equivalent, each having its costs and benefits. Both are also equivalent to exponential lags, which are a sort of combination of zero-order and first-order. However, with orders above first, both error and subjective workload increase dramatically. The reason that zero- and first-order systems are nearly equivalent may be appreciated by realizing that successful tracking requires that both position and velocity are matched (Poulton, 1974). In contrast, control systems of the second-order and higher are unequivocally worse than either zero- and first-order systems (Kelley, 1968).

The problems with second-order control are manifold: 1) one must anticipate its future state from its present, 2) the operator's effective time delay (response-lag) is also longer when higher derivatives must be processed and more computational work is demanded under second-order control. This increased lag contributes an additional penalty to performance. Second-order systems may be controlled by two strategies: 1) continuously and 2) "bang-bang" (double-impulse or time-optimal control). Here the operator perceives an error and reduces it in the minimum time possible with an open-loop "bang-bang" correction. A "bang-bang" correction means that a change of the control into one direction, is immediately followed by a change into the opposite direction. Because this double-impulse strategy reduces large errors in the shortest possible time, it is referred to as a form of "optimal control". While the double-impulse control eliminates the need for continuous perception of error derivatives of smooth and analog control, it does not necessarily reduce the total processing burden. With "bang-bang" control a more precise timing of the response is required and an accurate "internal model" of the state of the system must be maintained in working memory in order to apply the midcourse reversal at the appropriate moment. It will also produce high velocities, however, there are conditions when a lower-velocity "smooth-ride" is preferrable.

In some systems the order of control changes over time. This means an additional problem that the operator must also detect that the control dynamics have changed.

DISPLAYS

Certain tracking displays have been modified in order to make the tracking task easier and induce humans to control like differentiators. In the case of complex systems requiring high-order controls, some ways and means have been devised for relieving the operator of the need to perform the mental functions that otherwise would be required. One of these is aiding. The operational effect of aiding is to take over from the HO such operations as differentiation, integration and algebraic addition. However, aiding should be used selectively because its effectiviness depends upon a number of factors such as the nature of the input signal, the control order and whether the system is a pursuit or compensatory type.

Display augmentation. This is a form of operating aiding where the operator is informed, advised, instructed or told what to do. It serves to show the system condition relative to typical goals of operators. A problem is the right form of information relative to the control actions needed and the goals set. Poulton (1974) distinguishes 3 types of diplay augmentation: rate augmented, quickening and predictor displays. The last two forms of displays are also referred to as "historical" displays, because they indicate, by extrapolation, what is likely to happen if nothing is done.
A rate augmented display is in its simplest form an additional instrument showing rate, like the speedometer of an automobile.

A Predictor display uses, in effect, a fast-time model of the system to predict the future state of the system (or controlled variable) and display this state to the operator. Predictor displays offer particular advantages for complex control systems in which the operator needs to anticipate, such as with submarines, aircraft, spacecraft, vessel management and aircraft management. Experimental evidence shows a rather consistent enhancement of control performance with a predictor display (Dey, 1972; cited in McCormick and Sanders, 1982). An excellent discussion of this topic with regard to submarine depth control is given in Kelley (1968).

Quickening (Birmingham and Taylor, 1954) presents only a simple indicator of "quickened" tracking error. Like predictive displays, it indicates where the system is likely to be in the future if it is not controlled and it is most appropriate where the consequences of the operator's actions are not immediately reflected in the behavior of the system, but rather have a delayed effect, the delay frequently caused by the dynamics of the system, as in aircraft and submarines. Unlike the predictive display, it has no indication of the current error. This has the disadvantage that there are certainly times when you want to know where you are and not just where you will be. An advantage over predictive displays is that it contains just one element and so is more economical of space. It should also kept in mind that quickening does not have an appreciable advantage in very simple systems, or in systems where there is no delay in the system effect from the control action and where there is already immediate feedback of such system response. In order to provide the benefit of display economy without incurring the cost of an inaccurate picture of the present Gill et al. 1982) developed a pseudoquickened display. The presented symbol accurately corresponds to true position and error is indicated by intensity changes. The description of quickening by McCormick and Sanders (1982) is in fact a mixture of quickening and aiding in that the operator is shown what response to make.

Poulton raised serious questions (1974; p180-185) about research strategies used in some studies and also referred to certain disadvantages of quickening. In general he concluded that "true motion" predictor displays are likely to be far easier and safer to use for control systems of high order than quickened displays.

Preview of the input is of great value to an operator in a tracking task. The large benefits of preview result primarily because it enables the operator to compensate for processing lags in the tracking loop. Kvalseth (1979) indicates that such preview is most beneficial if the preview shows that portion of the track that immediately precede the "present" position. Duration of preview seems to be of less consequence than the opportunity to have at least some preview, but a preview span of approximately 0.5 sec seems to be mimimal (Kvalseth, 1978). The fact that the operator's time delay is in the order of 200-500 msec suggests that half a second of preview should be all that is needed when one is tracking a system that has no lags of its own (Reid and Drewell, 1972; cited by Wickens, 1984). With longer system lags more preview in the future is needed. In the absence of preview the operator must predict the future course of the system without perceptual guidance.

Anticipation refers to the operator's ability to predict what the future course will be without having any visible preview. Prediction is better if inputs have some systematic pattern and a low bandwidth. (see also the section on limits of the HO).

Pursuit displays generally provide superior performance to compensatory displays for two major reasons (Poulton, 1974). These relate to the ambiguity of compensatory information and the compatibility of pursuit displays. Ambiguity for the operator arises with compensatory displays because he is unable to distinguish between the three potential causes of error: command input, disturbance input and the operator's own control action. It will be obvious that pursuit displays by their nature are more compatible than compensatory displays.

Auditory displays The auditory modality is hampered somewhat because it does not have any precisely defined spatial reference points as vision does. Yet, under certain conditions auditory spatial displays can provide valuable supplementary information, particularly if the information is presented along channels that do not peripherally mask the comprehension of speech input. Since the auditory channel is more intrinsically tuned to processing verbal (speech) information, the use of auditory displays in tracking has received only minimal attention.

CONTROL COMPATIBILTY

As noted above, compatibility between input and output is a highly important aspect of tasks. Three types of compatibility are generally distinguished: spatial, movement and conceptual compatibility. Regarding spatial compatibility the physical similarity of the display and the controls and their physical arrangement are critical for the ease with which the control-display relationship is understood. The relationship between a movement of a control device and the movement on the display or by the system can have various forms. The control device and the display may differ in kind of movement, such as rotary versus linear, or in orientation i.e. the same vs. different planes etc., which all affect the compatibility of the movement relationships. Types and features of control devices and a number of principles of movement compatibility are described by McCormick and Sanders (Chapter 8 and 9; 1982) and Poulton (chapter 15; 1974). Conceptual compatibility relates to associations between coding systems, symbols, or other stimuli; these associations may be intrinsic or they may be culturally acquired.

LIMITS OF THE HUMAN OPERATOR

There are major limits affecting the operator's ability to perform a tracking task: limits in processing time, information-transmission rate, predictive capabilities, processing resources, and compatibility.

A certain processing time, commonly referred to as the effective time delay, is needed to translate a perceived error. Its absolute magnitude seems to depend somewhat upon the order of the system being controlled. Zero- and first-order systems are tracked with typical time delays from 150 to 300 msec. For a second-order system, the delay is longer, about 400 to 500 msec, reflecting the more complex decisions to be made.

When two input changes follow closely one after the other, the response to the second change is likely to be delayed. This reflects the psychological refractory period. Expectation of two changes in rapid succession may delay the response to the first one and make a combined response to both. There arises an interpretation problem when interpreting a double step response. In some cases it is difficult to discern between a preprogrammed double response output or a response consisting of two separate responses which run into each other (Poulton, 1974).

Time delays, whether the result of human processing or system lags are harmful to tracking for two reasons: (1) Obviously, any lag will have the effect that output no longer lines up with input. The error will increase with the magnitude of the delay. (2) Delays will induce problems of instability producing oscillatory behavior, when periodic or random inputs are tracked.

Limits of information transmission in tracking are between 4 and 10 bits/sec., depending upon the particular conditions of the display. The maximum transmission rate, for example, is considerably greater with pursuit than with compensatory displays (Crossman, 1960). When preview of input is available the transmission rate is also increased. The frequency rather than the complexity of making corrective decisions is more restrictive. The frequency limit in turn determines the maximum bandwidth of random inputs that can be tracked succesfully; it is normally found to be between 0.5 and 1.0 Hz. The maximum bandwidth can be increased to 2-3 Hz when the signals are predictable (Pew, 1974).

More serious limits than inputs at too high a bandwidth appear when operators track systems, like ships, that are characterized by lags. In this case the operator must make control corrections that will only be realized by the system output after a considerable time. In that case the corrective response requires anticipation i.e. prediction of future errors on the basis of the present values. Derivatives of the error signal, such as velocity and acceleration, must often be observed in tracking tasks. Humans perceive position changes more precisely than velocity or acceleration changes. Thus, anticipation will often fail to be precise in tracking tasks that demand perceptual systems to perform functions for which they are relatively ill-equipped.

In tracking it is generally assumed that operators continuously process the difference between where they are and where they would like to be and respond appropriately. In this way they compensate for error. This style of tracking is referred to as compensatory tracking. pursuit tracking behavior consists of only responding to input information and ignoring the output. In a sense the tracking responses are preprogrammed. Pursuit behavior leads to more efficient tracking because, unlike compensatory behavior, pursuit behavior does not require an error in order to generate a corrective response. Pew (1974) has reported that subjects are able to anticipate an upcoming input (pattern), by showing that the expected response was given although the input did not correspond to it.

In tracking tasks the operator must perform calculations and estimations of where the system will be in the future given an internal model of the system dynamics. These operations demand processing relating to a working memory. Because the processing resources of working memory are limited tracking is readily disrupted by concurrent tasks. The limits of human resources also account for tracking limitations when performing more than one tracking task at once, that is, in dual-axis tracking. For similar reasons, a self-paced tracking task is easier than one which is externally paced.

Because tracking is primarily a spatial task, it is apparent that compatibility relations affect performance. The research on control and display relations in tracking suggest that they do indeed.

MODELS OF THE HUMAN OPERATOR

The mathematical models of tracking performance that have been derived, have been some of the most accurate, succesful and useful of any of the models of human performance. Two models will be considered: the Crossover Model and the Optimal Control Model.

Crossover model. The early efforts of the late 1940's and 1950's for modeling the human operator were not very succesful. In these attempts one tried to discover the invariant characteristics of the HO as a transfer function relating perceived error and the response of the control device. The Crossover model, developed by McRuer and Krendel (1959), was more succesful because it looked for an invariant relationship between perceived error and the response of the system. In contrast to earlier attempts, the Crossover model allows the operator-describing function to be flexible and change with the system transfer function in order to achieve low error and a high degree of system stability. So, the model asserts that the HO responds in such a way as to make the "total" transfer function i.e. behavior of the HO and the System. The Crossover model considers the HO-system "team" as a first order system that can be described by two parameters: gain and effective time delay (HO response lag). The model is applicable to zero-, first- and second-order control dynamics, but not to third- and higher-order.

The Crossover model has proven to be quite succesful in accounting for human behavior in manual tracking. It has helped design engineers, it provided a useful means of predicting the mental workload encountered by aircraft pilots from the amount of lead or derivative control and it provided a convenient means of capturing the changes in the frequency domain that occur as a result of such factors as stress, fatigue, dual-task loading, practice or supplemental display cues.

However, the Crossover model has also its limitations. It is essentially a frequency-domain model, so it does not readily account for time-domain behavior. The model and its parameters are not derived from considerations of the processing mechanisms actually used by the HO. Unlike models of reaction time, signal detection, or dual-task performance, the Crossover model does not readily account for different operator strategies of performance.

The Optimal control model incorporates an explicit mechanism to account for strategic adjustments. A critical element of the Optimal control model is the quadratic cost functional which describes the trade-off between control precision and control effort. It assumed that the HO tries to minimize the outcome of this cost functional. Optimal control is not perfect control. The HO suffers two kinds of limitations: time-delay and disturbance for which he needs to engage into two extra processing operations: optimal prediction to compensate for the time-delay and estimation of the true state of the system from the noisy state.

Disadvantages of the optimal control model are the computational complexity as well as the greater number of parameters that must be specified to "fit" the data. These make it somewhat more difficult to apply. Nevertheless, the ability to account for shifts in operator strategies gives it a desirable degree of flexibility that the Crossover model does not possess. The model has been applied to optimize design of aviation systems, to assess operator workload and to assess attention allocation in a quantitative model of attention.

GENERAL ADVANTAGES AND DISADVANTAGES OF MATHEMATICAL MODELS

A number of advantages of mathematical descriptions of human behavior in dynamic control systems can be enumerated. They show the integration of a human and a machine working together in a way that performance measures can be predicted. In this way various design aids can be investigated for better or worse performance. Mathematical models may serve as a known reference point for describing relevant features of behavior. However, they are no substitute for an accurate desciption of behavior even when its results are precise.

A critical problem is that many of the manual control models, particularly the earlier ones, do not account for how humans filter, identify, and interpret potential information about them. Because of this inadequacy control models often predict identical performance regardless of the types of visual and auditory displays used, whereas human performance typically varies greatly with alternative forms of displays and display formatting of the same data. A second point of criticism is related to the first one. Models do not allow for effects of human memory of similar past situations. Third, human interpretation of previewed information is not fully achieved from the current derivatives of control conditions. Part of this interpretation is affected by the internal representation of the operating system that can only be vaguely mimicked by a mathematical model. Additional problem is that there are also times where people display shifts in criteria and behavioral discontinuities that are very difficult to model mathematically.

MULTIAXIS CONTROL

Humans must often perform more than one tracking task simultaneously. Even riding a bicycle involves tracking of lateral position while also stabilizing the vertical orientation of the bike. In general, there is a cost to multiaxis control that results from the division of processing resources between tasks. However, the severity of this cost is greatly influenced by the nature of the relation between the two (or more) variables that are controlled and the way in which they are physically configured.

A major distinction can be drawn between multiaxis systems in which the variables to be controlled as well as their inputs are essentially independent of each other, and those in which there is a cross-coupling so that the state of the system or variable of one axis partially constrains or determines the state of the other. Control of the heading and lateral position of an automobile are highly cross-coupled axes. Because lateral position cannot be changed independently of control of heading, the two cross-coupled tasks are also hierarchical. Many higher-order control systems in fact possess similar hierarchical relationships.

Multiaxis control is harmed if the error or output indicators are wider separated across the visual field. The obvious solution for the problem of display separation is to minimize this by bringing the displayed axes closer together. In this way peripheral interference is less of a cost to multiaxis tracking. Besides display separation other sources of diminished efficiency may be identified. Three such sources, relating to resource demand, control similarity and display-control integration will be considered.

Navon et al. (1982) note that the cost of multiaxis tracking with a single display and control is surprisingly small. As a general principle it may be stated that the cost of dual-axis control increases as the resource demands of a single axis are increased. Regarding the aspect of similarity of control dynamics it is well-known that when a single control strategy can be used for both axes simultaneously a better performance is achieved. Wickens, Tsang and Benel (1979) have found that the requirement of sharing different dynamics is also a contributor to increased subjective mental load, as well as reduced time-sharing efficiency. The display and control integration can be varied independently when two axes are tracked. Results of Chernikoff and Lemay (1963) show that when two axes with similar dynamics are shared, there is an advantage of integrating displays and controls and that the effect of integrating displays was generally more beneficial than that of integrating controls. In the former case, a clear reduction in visual scanning is produced, while in the latter case the possibility of response interference is increased. When different control strategies are required (competition), proximity should be minimized by separation of control and display. Because humans have problems in executing different independent responses - at least as long as they are not highly practiced - it is better to separate controls in the case of mixed dynamics.

MEASURES OF TRACKING: ERROR, REACTION TIME, INSTABILITY

Tracking error is defined as the deviation between the position of cursor and target. Error typically arises from one or two sources. Command inputs are changes in the target that must be tracked. For example, if the road curves, it will generate an error for a vehicle traveling in a straight line and so will necessitate a response. Disturbance inputs (noise) are those applied directly to the system. In the case of vehicle control a wind gust that buffets the car off the highway is a disturbance input. So also, is an inadvertent movement of the steering wheel by the driver. Either kind of input may be transient or continuous.

Tracking error is calculated at each point in time and then cumulated and averaged over the duration of the tracking task which lies generally between 30 sec and a few minutes. When tracking a moving object, engineers generally use the root-mean-square (RMS) error. RMS is calculated like the standard deviation of the error but without correcting each individual error value by the average constant error. It therefore includes a bias component, the average constant error, and a consistency component, the standard deviation. In the skill domain the RMS is often called the Total error. The rationale for excluding the constant error is that when tracking an irregular track that moves from side to side, the operator is as likely to be on one side of the correct position as on the other side, that is a constant error of zero.

When an operator follows a track that moves irregularly from side to side, he usually reproduces the position of the track, but with a time lag. Thus it is possible to measure the lag or lead error in time at each position, instead of measuring the more usual error in position at each time (see above). Engineers also analyze the operator's response by frequency. Here the average amplitude at each response frequency is plotted as the proportion of the average amplitude of this frequency in the track. The average lag in phase is also computed. Frequency methods of analyzing an operator's tracking performance are not usually much help in understanding what he is attemting to do, because the key questions cannot be rephrased in terms of frequency.

In connection with the frequency method the measure remnant is commonly used. The remnant is the part of the operator's response which does not correlate with the track. So, it is not represented in the transfer function. The remnant has three rather different sources. First the variability in phase, these are transient phase lags on either side of the average. Secondly, non-linear strategies used by the operator such as an onoff or bangbang strategy of control and thirdly variability caused by muscle tremor. The remnant is large when the operator adopts a non-linear strategy in an attempt to keep down his tracking error. This happens with tracks of high frequency (about 2.5 Hz) or higher-order control systems. The remnant is small when the operator can succesfully predict the track and preprogram his responses. This is the case with a track of low frequency and a position control system.

Frequency methods and methods of scoring which should not be used such as time-on-target (TOT) are further described by Poulton (1974; chapters 3 and 4). Kelley (1968) also gives a good discussion of different means of calculating tracking performance.

Reaction time measures are not commonly used in tracking of a continuous input. The reason is that the start of the response cannot be specified exactly, because the limb is already moving when the stimulus appears. If the time and direction of the stimulus are partly predictable, the operator will sometimes respond without waiting for the stimulus. When he predicts incorrectly, the stimulus for his correction may be the original stimulus, not the start of his incorrect response.

A major concern in the control of real-world dynamic systems is whether or not control will be stable, - that is whether the output will follow the input and eventually stabilize without excessive oscillations. Oscillatory and unstable bahavior can result from two quite different causes: positive and negative feedback. Positive feedback means that an error once in existence is magnified. Like second-order systems, those systems with positive feedback are universilly harmful for the obvious reason that they cannot be left unattended.

Systems with negative feedback function in such a way as to reduce detected errors. Instability caused by negative feedback results from a high gain coupled with large phase lags. A remedy is to reduce the gain. There is also an alternative strategy when the lag is long. Then one has to base the corrections on the trend of the error rather than its absolute current value.

CONCLUSIONS

A basic human factor design philosophy is that the human should be made to function as a zero-order controller when practical. More often, however, a low order of control, such as first-order (rate-/velocity-control) is the optimum choice. Long delays should also be avoided.

For situations in which a low order and a short lag of the control system cannot be realized, several means are available to relieve the task of the human operator. Showing HO what responses to make - that is, aiding, or telling where the system output will be on a predictor display improve tracking performance considerably in these cases. However, care should be taken HO not to overload with information; auxilliary information and instructions should selectively be applied.

Models of the human operator in tracking have neglected human information processing. Choosing what kind of information is the best for a particular tracking situation is not easily predicted. Future research should therefore be more directed to the psychological processes involved in a tracking task to enlarge the prediction power of tracking models.

REFERENCES

Birmingham, H.P. & Taylor, F.V. (1954). A design philosophy for man-machine control systems. Proceedings I.R.E., 42, 1748-1758.

Chernikoff, R., & Lemay, M. (1963). Effects of various display-control configurations on tracking with identical and different coordinate dynamics. Journal of Experimental Psychology, 66, 95-99.

Frost, G. (1972). Man-Machine dynamics. In H.P. van Cott and R.G. Kinkade (Eds.), Human engineering guide to equipment design. Washington D.C., American Institutes for research.

Gill, R., Wickens, C.D., Donchin, E., & Reid, R. (1982). Pseudo-quickening: A new display technique for the control of higher order systems. In R.E. Edwards (Ed.), Proceedings, 27th annual meeting of the Human Factors Society. Santa Monica, CA: Human Factors.

Keele, S.W., & Posner, M.I. (1968). Processing of visual feedback in rapid movements. Journal of Experimental Psychology, 77, 155-158.

Kelley, C.R. (1968). Manual and automatic control. New-York: J. Wiley and sons.

Kvalseth, T.O. (1978). Effects of preview on digital pursuit control performance. Human Factors, 20, 371-377.

McCormick, E.J., & Sanders, M.S. (1982). Human Factors in engineering and design. London: McGraw Hill.

McRuer, D.T., & Krendel, E.S. (1959). The human operator as a servo system element. Journal of the Franklin Institute, 267, 381-403, 511-536.

Navon, D., Gopher, D., Chillay, W. & Spitz, G. (1982). On separability and interference between tracking dimensions in dual axis tracking. Technical Report of the Institute of Technology. Haifa, Israel.

Pew, R.W. (1974). Human perceptual-motor performance. In B.H. Kantowitz (Ed.), Human Information Processing: Tutorials in performance and cognition. London: J. Wiley and sons.

Poulton, E.C. (1974). Tracking skill and manual control. London: Academic Press.

Poulton, E.C. (1981). Human manual control. In V.B. Brooks (Ed.) Handbook of Physiology: The nervous system, (Vol. 2). Bethesda: American Physiological Society. p. 1337-1389.

Scmidt, R.A. (1982). Motor control and learning. Champaign, Ill: Human Kinetics Publ.

Wickens, C.D., Tsang, P., & Benel, R. (1979). The dynamics of resource allocation. In C. Bensel (Ed.), Proceedings, 23rd annual meeting of the Human Factors Society. Santa Monica, CA: Human Factors.

Wickens, C.D. (1984). Engineering Psychology and human performance. Columbus: C.E. Merril Publ. Co.

5.4.2. TIME SHARING AND DUAL TASK PERFORMANCE
by Mike Donk

INTRODUCTION

This section aims to give an outline of the most important developments and issues within the resource theories of human performance. In the first part resource volume notions will be reviewed followed by a short examination of their underlying assumptions. A critical issue turns out to be the assumption of task invariance which refers to the necessity that tasks are not allowed to lose their independence when carried out together (Gopher & Sanders, 1984). The second part is concerned with the resource strategy view (Rabbitt, 1979) which, in contrast to the volume notions, is based on almost no assumptions at all. The last part of this section will present some recent theoretical developments followed by some concluding comments regarding future research needs in this area.

RESOURCE VOLUME THEORIES

In human performance theory, the quality of performance is often interpreted as the result of some basic limitation of the Human Processor. In resource theories the concept of processing resources is proposed to account for variations in efficiency with which tasks can be performed in combination. It is assumed that the organism possesses some kind of limited capacity (resources) needed to perform a task. Furthermore resources can eventually be distributed between tasks as well. The subject is presumed to be able to allocate capacity in different shares among concurrently performed tasks. When the joint resource demands of the tasks exceed the available capacity, performance on one or both tasks will decline. By overloading the organism -i.e. by forcing him to do more than he is able to manage at once within capacity limits- it becomes possible to investigate the volume of resources and the priorities of allocation.

This is exactly the idea behind the dual-task method of measuring "perceptual-motor load". This method was introduced originally by Bornemann (1942) with the intention to measure the "spare capacity" of a first task by means of a second task in order to determine the processing requirements of the primary task. In this way, the performance in the second task is assumed to become worser, the larger the resource consumption of the first task. Although the measurement of secondary-task performance as an index of resource expenditure has a high degree of face validity, the technique is not without problems.

In the first place it is necessary to protect the primary task against degradation when carried out together with the secondary task. This is often difficult to achieve, yet it is a necessary pre-equisite when interpreting the value of the secondary-task performance in relation to the capacity demands of the primary task.

Another important problem in measuring load through dual tasks concerns the observation that secondary task decrement is not always associated with greater processing requirements of the primary task (Kantowitz & Knight, 1974, 1976; Israel, Wickens, Chesney & Donchin, 1980; Whitaker, 1979). Such findings can partly be explained by non-resource-related factors such as structural interference -which relates to instances such as the difficulty in simultaneously performing two independent motor acts (e.g. rubbing the head and patting the stomach)- or peripheral interference -referring to the decrements in performance resulting from physical constraints (e.g. one can not say two words at the same time) (Kahneman, 1973; Wickens, 1984).

An additional and theoretically more relevant explanation has to do with the distinction between single vs. multiple resources. Are the two tasks tapping the same capacity or aren't they ? Finally the question may arise whether the tasks actually remain invariant when performed single or in combination. Perhaps they are integrated into a new "whole" rendering any combination unique and not comparable along a load scale (Rabbitt, 1979).

Prior to discussing some central themes of resource theories, a brief outline of the theoretical and methodological developments in this area will be presented.

THE SINGLE RESOURCE VIEW

Moray (1967) was among the first to propose that attention can be conceived of in terms of the limited processing capacity of a general purpose computer. This capacity could be allocated in graded amounts to various activities performed, depending upon their difficulty or demand for capacity (Moray, 1967). From its early beginning the capacity notion has emphasized the flexible and sharable nature of attention or processing resources.

During the 1970s the concept of capacity or resources as an intervening variable in dual task performance has been greatly elaborated by theoretical treatments of Kahneman (1973), Norman & Bobrow (1975) and Navon & Gopher (1979). Especially Norman & Bobrow (1975) contributed considerably to the development of resource theories by their introduction of the construct of the Performance Resource Function (PRF), an hypothetical function relating the quality of performance to the quantity of resources invested in a task.

It is assumed that the quality of performance is a monotonically nondecreasing function of the hypothetical resources invested in a task. Furthermore, an important distinction can be made between a data-limited region and a resource-limited region of the PRF-function. A task is said to be data-limited if the quality of performance does no longer depend on more or less resource investment. This can be caused by a very easy task in which performance can not be improved because the quality is already perfect, as well as by a very difficult task the performance of which cannot be improved irrespective of how hard one tries. In all other cases more or less resource-investment leads to a change in performance quality so that the function is resource-limited.

Because the PRF is only a hypothetical construct, it is almost impossible to derive the function from a dual task experiment; To construct a PRF reflecting the actual PRF it would be necessary that subjects allocate their resources as [illegible] and, in addition, the resources deployed in the two tasks [illegible] equivalent and maximal effective for each task. While the first condition can be met to a certain degree, the second condition [illegible] fulfilled.

By cross-plotting [illegible] under different priority conditions, a Performance Operating Characteristic (POC) (Norman

& Bobrow, 1975; Navon & Gopher, 1979) is obtained. This function describes optimal performance combinations depending on the distribution of the resources over the tasks. This has proven to be a useful tool in summarizing a number of characteristics of two time-shared tasks. Wickens (1984) has distinguished some important characteristics of this function. First, single task performance is represented on the axes of the POC. When single task performance is better than dual task performance with absolute emphasis on one task, a cost of concurrence is observed. This can be the result of an extra resource demand of an executive time-sharer that is utilized only in a dual task condition (Hunt & Lansman, 1981; Moray, 1967; McLeod, 1977; Taylor, Lindsay & Forbes, 1967); in addition extra costs can be due to peripheral or structural interference. Second, time sharing efficiency is the effectivity with which two tasks can be done at once. This efficiency is high when there is almost no decrement in performance of each individual task when they are performed together and low when large performance detoriations are observed as a result of time sharing. Third, the degree of exchange indicates the extent resources are shared or exchangable between tasks. A distinction is made between a rectangular POC that is essentially without any exchange and a more smooth POC in which some degree of exchange is always present. Fourth, the allocation bias is indicated by the proximity of a certain point on he POC to one axis in comparison with the other. This bias is presumed to be determined by the resource allocation between the two tasks.

Although the POC-method relies on several strong assumptions, empirical POCs certainly provide a fruitful summary of the nature of the underlying PRFs as well as of various apparently different phenomena; Variables like task difficulty, amount of practice received by the subjects, automatic vs. controlled processing (Shiffrin & Schneider, 1977), parallel vs. serial processing, are elegantly described by the same basic argument of resource theory. The only way they differ concerns the varying extent to which resource investment can bring about changes in performance equality. This implies that the easier a task is, the more practice has been received, or the more automatically information is processed, the less resource investment is required to perform the task. An easy or practiced task has a larger data-limited region or less resources are required to bring about a change in performance. Although such assumptions look straightforward, they have wide theoretical implications. For example it may not be useful to speak of dichotomies such as controlled vs. automatic processes since their only difference is concerned with quantitative resource requirements. In the same way the basic distinction disappears between easy and difficult tasks or between strategical changes in performance with practice. Logically, low-resource demanding tasks--as a result of practice or ease--can be done in parallel while, on the contrary, difficult or attention demanding tasks have to be serially performed because their demand for resources per unit of time would otherwise exceed the available capacity. An example will make this point clear.

In one experiment Schneider & Fisk (1982) have examined the ability of combining automatic and controlled processing. The subjects had to perform a consistent mapping (CM) search -letters among digits- on one diagonal and a varied mapping (VM) search -digits among digits- on the other diagonal of a visual display (Shiffrin & Schneider, 1977; Schneider & Shiffrin, 1977). The CM search is supposed to be completely automatic due to extensive practice subjects have received during their lives to distinguish letters from digits. The VM search requires controlled processing implicating that performance in this task depends on resource investment i.e. performance is resource limited in this task.

In another study they looked at the ability to perform two controlled tasks at the same time; two VM search tasks had to be performed in combination.The results showed that subjects could well perform a controlled task in combination with an automatic one, but two controlled attention-demanding and resource-consuming tasks interfered to a considerable extent. In a resource volume framework such results are well understood. A controlled task consumes a large amount of resources so that when two tasks are performed in combination their resource requirements exceed the available capacity and a drop in performance can be observed. When a controlled task is performed in combination with an automatic one, the whole capacity can be allocated to the controlled task, and both tasks can be easily done in parallel without loss of efficiency. The resulting POC has a rectangular form, which means that there is no resource trade off between the tasks.

PROBLEMS WITH AN UNDIFFERENCIATED SINGLE RESOURCE VIEW

The original capacity notion assumed a single reservoir of undifferentiated resources. However, a number of experiments suggests that this view is too simple. First, there is the finding that some tasks can be time-shared without considerable loss of efficiency in either task (Allport, Antonis & Reynolds, 1972; Wickens, 1976; Shaffer, 1975; Kleiman, 1975; Rollins & Hendricks, 1980; Treisman & Davies, 1973). Different experiments demonstrate that when two tasks differ sufficiently from one another, they can be performed in combination. In this way it is reported for example that subjects could concurrently sight-read music and engage in an auditory shadowing task as well as they could perform either task alone (Allport, Antonis & Reynolds, 1972). Such results could eventually be explained by a single capacity theory assuming at least a moderate level of automaticity or the presence of data-limited regions in one of the tasks. In all fairness, however, this is probably not the case in view of the usually unpredictable nature of the stimuli and the relatively heavy time pressure.

A second problem for the single capacity hypothesis stems from experiments in which a change in difficulty of the primary task fails to influence the performance of the secondary task although the performance of the primary task remains the same (North, 1977; Isreal, Chesney, Wickens & Donchin, 1980; Kantowitz & Knight, 1976; Wickens & Kessel, 1979). In a study by North (1977) subjects had to time-share a tracking task with a discrete digit-processing task, which was varied in difficulty. Although the difficulty manipulation of the digit-processing task interfered considerably with an additional digit cancelling task, no effect was found on simultaneous tracking. Such a phenomenon is called "difficulty insensitivity" (Wickens, 1980) and implies that there is no difficulty-performance trade off. A single capacity notion could only survive by assuming substantial datalimits in the secondary task. However, from several experiments in which tracking was paired with other tasks, the conclusion is that this is not the case with tracking (Wewerwinke, 1976; Briggs, Peters, Fisher, 1972; Johnston et al., 1970; Shulman & Briggs, 1971; Watson, 1972; Cliff, 1973; Danos & Wickens, 1977; Glucksberg, 1963).

A third difficulty concerns "structural alteration effects", which can often be observed when two tasks are time shared and one of them has a change in processing structure, such as input or output modality or memory code. Under such circumstances a change in interference between the two tasks has been repeatedly observed although as such the difficulty of both tasks remained unaltered (Isreal, 1980; Martin, 1980; Rollins & Hendricks, 1980; Treisman & Davies, 1973; Vidulich & Wickens, 1981; Wickens et al., 1983; Harris, Owens & North, 1978; McLeod, 1977; Wickens, 1980; Friedman, Polson, Dafoe & Gaskill, 1982; McFarland & Ashton, 1978; Wickens & Sandry, 1982; Pritchard & Hendrickson, 1985). It is certainly impossible to account for such results by a simple single capacity notion.

Fourth, the phenomenon of "uncoupling of difficulty" can be mentioned. This refers to instances in which, when paired with a third task, the more difficult one of two tasks actually interferes less with the third task than the easier one (Wickens, 1976, 1980). This last observation is also incompatible with the single resource concept.

It can be said that the most important shortcoming of single capacity theory concerns the neglect of the structural aspects of the tasks which, in contrast, have been overemphasized by the structural theories (Broadbent, 1958; Treisman, 1960; Deutsch & Deutsch, 1963). Resource volume theory has to account for these structural aspects in order to provide a proper explanation for the above-mentioned results. There are three possible alternatives to the single resource view;

A first possibility is to adhere to a single capacity view and to assume additional auxiliary structures (Kahneman, 1973). In this way tasks have to compete for a general pool of resources (effort) as well as for more or less dedicated satelite structures (e.g. modalities). Although the model of Kahneman (1973) can explain the results that gave difficulties for the original single capacity notion, it remains rather vague about the precise nature of the satelite structures. Every result could be accounted for by assuming a new structure, undermining in this way strongly the predictive power of the model. A second alternative is to consider multiple resources which are at least to some extent interchangable between tasks (Navon & Gopher, 1979; Wickens, 1980; Sanders, 1979). Third, there is the strong assumption of task invariance which could lead to dropping volume and adopting a resource strategy model (Rabbitt, 1979; Hockey, 1979).

In the next part a more elaborated review of the multiple resource notions will be given followed by a brief summary of the strategy view. The last part is involved with some recent developments in favour of and against resource notions. It will conclude with some suggestions for future research.

MULTIPLE RESOURCE VOLUME THEORIES

A second alternative to single capacity theory supposes the existence of multiple resources (Friedman et al., 1982; Kantowitz & Knight, 1976; Navon & Gopher, 1979; Sanders, 1979; Wickens, 1980; Kinsbourne & Hicks, 1978). According to this multiple resource theory the Human Processor possesses more than one commodity with resource-like properties such as allocation, flexibility and sharing. These resources can be allocated in graded amounts only within the structures they relate to. A distinction can be made between multiple resources residing in separate structures

(Kinsbourne & Hicks, 1978; Baddeley & Lieberman, 1980; Navon & Gopher, 1979; Wickens, 1980) and multiple resources that are beyond structures (Sanders, 1983). In the first type the emphasis is on competition for resources; in the second type there is competition for structures as well as for resources. By making a distinction between resources and structures, the second approach has some advantages above the original multiple resource concept in which resources are defined by structures. For the present we will restrict ourselves to the first multiple resource view and will return to the other type later on.

The assertion that more than one resource is involved in information processing is quite well accepted. However, establishing the identity of the specific resources constitutes a considerable difficulty. In the first place it is necessary to achieve at least some degree of parsimony in the number of proposed resource-systems, otherwise the entire concept of structure-specific resources will rapidly lose predictive and explanatory power and ultimately share the fate of classical "faculty" notions. This means that no new resource can be defined for each task element. The experimental results suggest a distinction between resources concerned with processing stages, resources relating to hemispheric processing and resources that concern modalities of processing:

- Processing Stages: A number of experiments have provided evidence that tasks that primarily rely on perceptual processing can efficiently be time-shared with tasks whose demands are primarily response related (Trumbo, Noble & Swink, 1967; Wickens, 1976). In contrast, two perceptual or two response loading tasks interfere to a considerable extent (Long, 1976; Treisman & Davies, 1973). Furthermore also the phenomenon of "difficulty insensitivity" (Wickens, 1980) is often shown in experiments in which two tasks are used which seem to rely on different processing stages (Isreal, Chesney, Wickens & Donchin, 1980; Kantowitz & Knight, 1976; Roediger et al., 1977; Wickens, Isreal & Donchin, 1977; Wickens & Kessel, 1979).

- Hemispheres of Processing: With reference to resources relating to hemispheric processing, evidence is provided by research of Kinsbourne & Hicks (1978) who observed larger interference when a verbal task was combined with a second task in which the right hand -corresponding to the left verbal hemisphere- was involved than with one in which this was the left hand -corresponding to the right spatial hemisphere-. Brooks (1968) also showed this hemispheric specificity, even within one task; a task requiring spatial working memory was performed better in combination with a verbal response while a task relying on verbal working memory could better be performed with a spatial response. Furthermore, reaction time is lengthened when the hemisphere of stimulus processing is the same as that controlling the responses (Allwitt, 1981; Dimond & Beaumont, 1972; Wickens & Sandry, 1982).

- Modalities of Processing: The last dimension is a more difficult one to establish because of the somewhat conflicting results. Some studies suggest that there is indeed an advantage by cross-modulating two tasks (Harris, Owens & North, 1978; McLeod, 1977, 1978; Glucksberg, 1963; Treisman & Davies, 1973; Wewerwinke, 1976) while others do not find this advantage (Lindsay, Taylor & Forbes, 1968; Trumbo & Milone, 1971).

Despite the somewhat doubtful state of modality specific resources, Wickens (1980) has integrated all three dimensions in one 3-dimensional-cube model. While this model defines concrete resources and adheres also to the demand for parsimony, it provides a framework that can be tested. However, the ultimate test is not easy.

The second difficulty in identifying multiple resources concerns the use of the POC technique, especially its interpretation. Most resource theorists (Norman & Bobrow, 1975; Navon & Gopher, 1979, 1980; Gopher & Sanders, 1984; Wickens, 1980, 1984) consider the POC a valuable tool in describing dual task performance in terms of resource notions. Yet results obtained in dual task experiments can be explained by different causes. The only proper way to conduct dual task experiments with the intention of investigating wether more than one resource is involved, is by manipulating the difficulty as well as the priority of resource allocation (Navon & Gopher, 1979, 1980) under the assumption that all other subject-task parameters remain constant (Gopher & Sanders, 1984).

When two tasks share one resource a clear trade off has to be present in the POC. Yet, a smoothlike POC-form does not necessarily imply the sharing of one resource i.e. concurrence costs could occur as a result of structural or peripheral interference yielding a smooth POC although the tasks may be tapping different resources.

When two tasks demand different resources a rectangular POC-form has to be observed, yet, also in this case no guarantee is given wether the tasks tap different resources. In the case of large data-limits it could quite well be possible to become a boxlike POC although the tasks tap the same resource.

In conclusion, in a good experiment the possibilty of structural and peripheral interference has to be ruled out by choosing appropriate task-combinations. Furthermore, it must be reasonable to assume no data limits in either task. To be sure that this is indeed the case each individual task has to show interference with at least one other task with which it has been paired in advance. Accepting the hypothesis of multiple resources is only justified when these considerations are taken into account.

Gopher & Sanders (1984) have systematically discussed the assumptions that are necessary for allowing a POC interpretation;

First, resource allocation has to be invariant and maximal; this means that subjects have to dedicate their resources fully to performance and the available resource volume has to be fixed. If this is not the case the behavioral measures of the task performance will become basically unreliable in revealing something about the nature of the resources; the only alternative would be an independent psychophysiological measure of resource allocation (Kahneman, 1973).

A second assumption relates to the claim of Norman & Bobrow (1975) that performance has to be a monotonic nondecreasing function of resource investment. If this assumption does not hold, any interpretation of a POC in terms of the resource volume theories becomes useless.

Third, it is necessary that, at least to some extent, subjects can manage and allocate their resources, which is imperative for constructing a POC.

The fourth and probably most critical assumption concerns process or task invariance; this means that the two tasks are not allowed to loose their independence when carried out together and that subjects do not change their basic strategies in performing each individual task when task variables are manipulated i.e. different dual task priority combinations only reflect changes in the amounts of allocated resources (Gopher & Sanders, 1984).

This last assumption is not only the most important condition required for interpreting a POC but also the most doubtful one. Various experiments have shown that task-integration occurs under dual task conditions (Neisser, 1976; Hirst, Spelke, Reaves, Coharack & Neisser, 1980; Spelke, Hirst & Neisser, 1976; Lucas & Bub, 1981; Neisser, Hirst & Spelke, 1981). In several studies subjects were trained for many months to pick up two verbal messages -one visual and one auditory- at the same time (Hirst et al., 1980; Spelke et al., 1976). The results showed impressive practice effects although neither task had been processed at an "automatic" level. Results like these are very hard to combine with the assumption of task invariance. More extreme, they can be interpreted as a confirmation for the "attention-is-a-skill" hypothesis (Hirst et al.,1980; Spelke et al., 1976), which proposes that by way of task-integration extended practice in time-sharing suffices to eliminate dual-task interference.

In several other investigations concurrence benefits have been demonstrated (Johnson & McClelland, 1974; Reicher, 1969; Pomerantz, Sager & Stoever, 1977; Rabbitt, 1979). This implies that by pairing two tasks the performance on each one becomes better relative to single task performance. It has been known since long that, for example, a short familiar word can be better perceived than each individual letter on its own (Cattell,1885). More recently this superiority effect has been demonstrated for objects as well (Weisstein & Harris, 1974; McClelland, 1978; Wandmacher, 1981). The instances of concurrence benefits are obviously hard to reconcile with the notion of task invariance.

Results like these led Rabbitt (1979) and others to reject resource-volume notions in favour of a resource strategy theory in which qualitative strategical shifts in performance are emphasized and consequently the resource-driven nature of human information processing. This is, however, a drastic alternative to the resource volume notions and not without objections.

RESOURCE STRATEGY THEORY

The resource-strategy model does not assume invariance of the nature of the operations. Instead, the operations may undergo fundamental changes as a function of practice (Bainbridge, 1978), processing priorities or information load (Sperandio, 1972; Rabbitt, 1979). Basic to the resource strategy theories is that there occur qualitative changes in processing as a function of strategy. Within this framework the term "resource" is a vague concept referring to almost any processing capability, energetic as well as structural (Sanders, 1983). Resources are "acquired information about the structure of particular tasks and about the external world which are used by the subject in order to actively control their momentary perceptual selectivity and their choice of responses" (Rabbitt, 1979). Rabbitt (1979) emphasizes the active, top-down control of the Human Processor in performance. Furthermore, the locus of control within the human system can vary from time to time during a task depending on task-demands and the systems' idiosyncratic characteristics (Hamilton, Hockey & Reyman, 1977; Hockey, 1979; Rabbitt, 1979). Within the strategy notions, the most important research method is also the dual task paradigm but the main focus is on qualitative changes like neglecting peripheral elements (Bartlett,1953) or changes in allocation priorities. Furthermore it is important to realize that momentary selectivity and choice of responses is in this model fully determined by top-down control.

Although strategy models are quite popular, the framework is almost without predictive power because of a lack of assumptions. Once one allows qualitative changes when executing a task, any result can fit the model as another "qualitative change" and, consequently, formal theory building is almost impossible (Gopher & Sanders, 1984).

RECENT DEVELOPMENTS PRO AND CONTRA RESOURCE NOTIONS

Although multiple resource notions (Sanders, 1979; Wickens, 1980; Navon & Gopher, 1979) are still quite current, a structural interference view of dual task performance has been proposed as a viable alternative. Thus, Navon (1984) has discussed two fundamental criticisms with respect to the notion of resource volume. In the first place behavioral phenomena such as the quality of performance under different conditions, that invoke the introduction of the resource concept, can be equally well accounted for by intervening variables such as motivation and task difficulty. Furthermore, he states that performances that are of interest to cognitive psychologists, do not have serious constraints resulting from the scarce availability of energy supplies. Thus, Navon (1984, 1985) strongly questions the explanatory value of the resource concept in the interpretation of performance variability in such tasks as decision making, memory search, or interference between tasks in concurrent performance.

In a later article, Navon & Miller (1986) have proposed a view of dual task interference in which they stress the role of outcome conflict; The effects brought about by one task could change the state of some variable that is relevant for performing the concurrent task. In this view, outcome conflict is cross-talk among parallel processing lines. The main way in which a person attempts to overcome this conflict is by adopting a strategy of handling the tasks more sequentially. Furthermore, extended practice may change the way the tasks are carried out and reduce the amount of cross-talk.

In some aspects this view is similar to Rabbitts` strategy notion. Yet, the reasons for strategical control in avoiding outcome conflict are somewhat more pronounced which renders the theory more testable. Furthermore Navon also emphasizes the structural data-driven aspects of information processing through a late selection view in which the selective "filter" consists of a strategy, adopted by the subject, enabling sequential access to processing mechanisms which are subject to conflict.

A similar view is proposed by Neumann (1985), who suggests that the limits of attention are not due to processing limitations in the sense of limited capacity, but rather result from the way in which the brain solves selection problems in the control of action. He emphasizes that the difficulty of time sharing is not to combine stimuli, but rather to deal with them independently at the same time; selection is needed for the control of action (Neumann, 1986).

His criticism on resource volume theories is that they cannot account for all dual task results; interference is usually more specific as would be predicted on the basis of a limited number of resources; furthermore there are also cases of unspecific interference (Neumann, 1985). However his main criticism is that the resource volume notions provide no explanation why capacity is limited. They are limited to the statement that capacity is limited.

In line with the traditional structural late selection view as well as with theorists like Neisser (1976) and Allport (1980), Neumann states that an attention mechanism is necessary to avoid behavioral chaos that would result from an attempt to simultaneously perform all possible actions for which sufficient causes - e.g. motives, skills, appropriate stimuli- exist. This selection mechanism has to select skills and make them available in order to attain well stated action goals. He distinguishes two selection problems that can be encountered; First, the problem of effector recruitment -which skills, related to the goals of action, are given access to the effector system ?- and, second, the problem of parameter specification -which of the possible specifications of an action`s parameters is put into effect ?- (Neumann, 1986). To solve these two selection problems, attention mechanisms are a necessity to achieve proper performance.

The notions of Neumann are certainly important but a more precise specification of the properties of the attention mechanisms is needed to avoid the fate of post-hoc theorising that can explain anything but has no predictive power.

In summary, it can be said that these recent developments are characterized by strong objections to the resource volume concept. A more strategical view is proposed in which top-down processes as well as structural bottom up processes are important. Although these developments might provide fruitful insights in human performance, especially through the introduction of the "functionality question" (why is capacity limited ?), they are, as yet, not more than a first small step towards a model of attention.

The alternative approach to human attention remains in terms of resource volume. The fact that violations of the underlying assumptions about the interpretation of the POC are observed or may even be common, does not reduce the importance of spelling out the assumptions and identifying the nature of the violations and the instances at which they occur. The robustness of results obtained in several experimental situations using the same variables may enable one to assign proper weights to the consequences of different assumptions (Gopher & Sanders, 1984).

In addition, the attacks of Navon (1984) on volume notions are not as compelling as they may be thought at first sight. First, he states that the resource concept is only meaningful when considered as an intervening variable. According to Gopher (1985) this is not necessarily true; resources can be also conceived of as hypothetical constructs which are useful and productive for theory and research. Navon remarks that energy-limited considerations are irrelevant in most tasks of interest to cognitive psychologists. Although it can be argued that there are many tasks in which performance is not directly limited by resources there are other conditions in which energy allocation plays a prominent role; There is a continuum from short term tasks, in which resource considerations are minor to sustained attention tasks, within which the role of energy modulation is well accepted. What is needed is one framework that can account for energy considerations accross the whole domain of tasks.

Such a framework is offered by Sanders` energetic stage model (1983) in which an attempt is made to incorporate energetical concepts in stage thinking. In this model structural as well as the energetical aspects of information processing are included. In fact the model is not based on resource notions and its accompanying dual task methodology. It is a stage model in which the energetic dimension is tested by considering unusual circumstances such as sleep loss, drugs or noise. Furthermore, it contrasts sharply with the resource strategy notions in that it is based on the strict constraints of the stage logic (Sternberg, 1969).

SUMMARY

The critical assumption of task invariance in dual tasks is probably the most debatable notion of resource theory. Are tasks really remaining independent when they are carried out together or are they merely integrated into a new "whole" ?

From the foregoing discussion it appeared that in certain conditions task integration is observed (Neisser, 1976; Hirst, Spelke, Reaves, Coharack & Neisser, 1980; Spelke, Hirst & Neisser, 1976; Lucas & Bub, 1981; Neisser, Hirst & Spelke, 1981). An important issue is the question to what extent and under which conditions task invariance is a reasonable assumption and under which conditions is it not. More concretely; is it difficult for subjects to combine task elements in dual task performance or is it difficult to process them without combining them?

If task integration is the general phenomenon this would imply that the interpretation of human information processing in the sense of resource volume notions is wrong. In contrast, the strategical and more recent structural theories assume that the problem in dual task performance is even to keep processing two stimuli apart; In case of coordinated performance the Human Processor normally combines stimuli in order to attain well-stated action goals (Neumann, 1985) or to avoid confusion and cross-talk (Navon, 1985, 1986). The complication of the latter type of view is that, as yet, it does not spell out how and with which variety task integration may occur. Resource strategy theory merely states a top-down principle but do not describe or predict performance.

Thus, before interpreting results in terms of volume notions or strategy notions, the issue of task invariance should be more widely examined. Its outcomes are decisive for the future directions of the area.

REFERENCES

Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.), Cognitive psychology - new directions. London: Routledge & Kegan Paul.

Allport, D. A:, Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. Quarterly Journal of Experimental Psychology, 24, 225-235.

Allwitt, L. F. (1981). Two neural mechanisms related to modes of selective attention. Journal of Experimental Psychology: Human Perception and Performance, 7, 324-332.

Baddeley, A. D., & Lieberman, K. (1980). Spatial working memory and imagery mnemonics. In R. Nickerson (Ed.), Attention and performance VIII. Englewood Cliffs, N. J.: Erlbaum.

Bainbridge, L. (1978). Forgotten alternatives in skill and workload. Ergonomics, 21, 169-185.

Bartlett, F. C. (1953). Psychological criteria of fatigue. In W. F. Floyd & A. T. Welford (Eds.), Symposium on fatigue. London: Lewis.

Bornemann, E. (1942). Untersuchungen über den Grad der geistigen Beanspruchung. Arbeitspsychologie, 12, 142-191.

Briggs, G., Peters, G., & Fisher, R. P. (1972). On the locus of divided attention effects. Perception and Psychophysics, 11, 315-320.

Broadbent, D. (1958). Perception and communication. Oxford: Permagon.

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. Canadian Journal of Psychology, 22, 349-368.

Cattell, J. McK. (1885). The inertia of the eye and brain. Brain, 8, 295-312.

Cliff, R. C. (1973). Attention sharing in the performance of a dynamic dual task. IEEE Transactions on Systems Man and Cybernetics, SMC-3, 241-248.

Damos, D. L., & Wickens, C. D. (1977). Dual task performance and Hick`s law. Journal of Motor Behavior, 9, 209-215.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. Psychological Review, 70, 80-90.

Dimond, S. J., & Beaumont, J. G. (1972). Processing in perceptual integration between and within the cerebral hemispheres. British Journal of Psychology, 63, 509-514.

Friedman, A., Polson, M. C., Gaskill, S. J., & Dafoe, C. G. (1982). Competition for left hemisphere resources: Right hemisphere superiority at abstract verbal information processing. Journal of Experimental Psychology: Human Perception Performance, 7, 1031-1051.

Glucksberg, S. (1963). Rotary pursuity tracking with divided attention. Journal of Engineering Psychology, 2, 119-125.

Gopher, D. (1985). In defense of resources: On structures, energies, pools and the allocation of attention.

Gopher, D., & Sanders, A. F. (1984). "S-Oh-R": Oh stages! Oh resources! In W. Prinz, & A. F. Sanders (Eds.), Cognition and motor processes (pp. 231-253). Berlin etc.: Springer.

Hamilton, P., Hockey, G. R. J., & Reyman, R. (1977). The place of the concept of activation in human information processing. In S. Dornic (Ed.), Attention and performance VI. Hillsdale, NJ: Erlbaum.

Harris, S., Owens, J., & North, R. A. (1978). A system for the assessment of human performance in concurrent verbal and manual control tasks. Behavior Research Methods & Instrumentation, 10, 329-333.

Hirst, W., Spelke, E. S., Reaves, C. C., Coharack, G., & Neisser, U. (1980). Dividing attention without alternation or automaticity. Journal of Experimental Psychology: General, 109, 98-117.

Hockey, R. (1979). Stress and the cognitive components of skilled performance. In V. Hamilton, & D. M. Warburton (Eds.), Human stress and cognition (pp: 141-177). New York: Wiley.

Hunt, E., & Lansman, M. (1981). Individual differences in attention. In R. Sternberg (Ed.), Advances in the psychology of intelligence (Vol. 1). Hillsdale, N. J.: Erlbaum.

Isreal, J. B. (1980). Structural interference in dual task performance: Behavioral and electrophysiological data. Unpublished doctoral dissertation, University of Illinois.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 17, 259-273.

Isreal, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. Human Factors, 22, 211-224.

Johnson, J. C., McClelland, J. L. (1974). Perception of letters in words: Seek not and ye shall find. Science, 184, 1192-1194.

Johnston, W., Greenberg, S., Fisher, R., & Martin, D. (1970). Divided attention: A vehicle for monitoring memory processes. Journal of Experimental Psychology, 83, 164-171.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice Hall.

Kantowitz, B. H., & Knight, J. L. (1974). Testing tapping and time-sharing. Journal of Experimental Psychology, 103, 331-336.

Kantowitz, B. H., & Knight, J. L. (1976). Testing tapping and time-sharing II: Use of auditory secondary task. Acta Psychologica, 40, 343-362.

Kinsbourne, M., & Hicks, R. (1978). Functional cerebral space. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N. J.: Erlbaum.

Kleiman, G. M. (1975). Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 14, 323-389.

Lindsay, P. H., Taylor, M. M., & Forbes, S. M. (1968). Attention and multidimensional discrimination. Perception & Psychophysics, 4, 113-117.

Long, J. (1976). Division of attention between non-verbal signals: All or none or shared processing. Quarterly Journal of Experimental Psychology, 28, 47-69.

Lucas, M., & Bub, D. (1981). Can practice result in the ability to divide attention between two complex language tasks? Comment on Hirst et al.. Journal of Experimental Psychology: General, 110, 495-498.

McClelland, J. L. (1978). Perception and masking of wholes and parts. Journal of Experimental Psychology: Human Perception and Performance, 4, 210-223.

McFarland, K., & Ashton, R. (1978). The influence of concurrent task difficulty on manual performance. Neurophysiologica, 16, 735-741.

McLeod, P. (1977). A dual task response modality effect: Support for multi-processor models of attention. Quarterly Journal of Experimental Psychology, 29, 651-667.

Martin, M. (1980). Attention to words in different modalities: Four channel presentation with physical and semantic selection. Acta Psychologica, 44, 99-115.

Moray, N. (1967). Where is attention limited. A survey and a model. Acta Psychologica, 27, 84-92.

Navon, D. (1984). Resources- A theoretical soup stone? Psychological Review, 91, 216-234.

Navon, D. (1985). Attention division or attention sharing. In M. I. Posner & O. S. M. Marin (Eds.), Attention and performance XI. Hillsdale, NJ: Erlbaum.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. Psychological Review, 86, 214-255.

Navon, D., & Gopher, D. (1980). Interpretations of task difficulty. In R. Nickerson (Ed.), Attention and performance VIII. Hillsdale, N. J.: Erlbaum.

Navon, D., & Miller, J. (1986). The role of outcome conflict in dual-task interference (ICS Report 8601). California, San Diego: Institute for cognitive science.

Neisser, U. (1976). Cognition and reality. San Francisco: Freeman.

Neisser, U., Hirst, W., & Spelke, E. S. (1981). Limited capacity theories and the notion of automaticity: Reply to Lucas and Bub. Journal of Experimental Psychology: General, 110,499-500.

Neumann, O. (1986). ... In: H.Heuer, & A.F. Sanders, Issues in cognition and motor behavior. Hillsdale, NJ. (in press).

Norman, D., & Bobrow, D. (1975). On data limited and resource limited processing. Journal of Cognitive Psychology, 7, 44-60.

North, R. A. (1977). Task components and demands as factors in dual-task performance (Report No. ARL-77-2/AFOSE-77-2). Urbana, Ill.: University of Illinois at Urbana- Champaign, Aviation Research Laboratory.

Pomerantz, J. R., Sager, L. C., & Stoever, R. (1977). Perception of words and their component parts: Some configural superiority effects. Journal of Experimental Psychology: Human Perception an Performance, 3, 422-435.

Pritchard, W. S., & Hendrickson, R. (1985). The structure of human attention: Evidence for separate spatial and verbal resource pools. Bulletin of the Psychonomic Society, 23(3), 177-180.

Rabbitt, P. M. A. (1979). Current paradigms and models in human information processing. In V. Hamilton, & D. M. Warburton

(Eds.), Human stress and cognition (pp. 115-140). New York: Wiley.
Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 81, 274-280.
Roediger, H. L., Knight, J. L., & Kantowitz, B. H. (1977). Inferring decay in short term memory: The issue of capacity. Memory & Cognition, 5, 167-176.
Rollins, H. A., & Hendricks, R. (1980). Processing of words presented simultaneously to eye and ear. Journal of Experimental Psychology: Human Perception and Performance, 6, 99-109.
Sanders, A. F. (1979). Some remarks on mental load. In N. Moray (Ed.), Mental workload: Its theory and measurement. New York: Plenum Press.
Sanders, A. F. (1983). Towards a model of stress and human performance. Acta Psychologica, 53, 61-97.
Schneider, W., & Fisk, A. D. (1982). Concurrent automatic and controlled visual search: Can processing occur without resource cost? Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 261-278.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84, 1-66.
Shaffer, H. L. (1975). Multiple attention in continuous verbal tasks. In P. M. A. Rabbitt, & S. Dornic (Eds.), Attention and Performance V. London, Academic Press.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.
Shulman, H. G., & Briggs, G. (1971). Studies of performance in complex aircrew tasks. Ohio State University Research Foundation. Air Force Project 2718, Final Report.
Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. Cognition, 4, 215-230.
Sperandio, J. C. (1972). Charge de travail et regulation des processes operatoires. Travail Humaine, 35, 85-98.
Sternberg, S. (1969). On the discovery of processing stages: Some extensions of Donders` method. Acta Psychologica, 30, 276-315.
Taylor, M. M., Lindsay, P. M., & Forbes, S. M. (1967). Quantification of shared capacity processing in auditory and visual discrimination. Acta Psychologica, 27, 223-229.
Treisman, A. M. (1960). Contextual cues in selective listening. Quarterly Journal of Experimental Psychology, 12, 242-248.
Treisman, A. M., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), Attention and performance IV. New York: Academic Press.
Trumbo, D., Milone, F. (1971). Primary task performance as a function of encoding, retention, and recall in a secondary task. Journal of Experimental Psychology, 91, 273-279.
Trumbo, D., Noble, M., & Swink, J. (1967). Secondary task interference in the performance of tracking tasks. Journal of

Experimental Psycholöogy, 73, 232-240.

Vidulich, M., & Wickens, C. D. (1981). Time-sharing manual control and memory search: The joint effects of input and output modality competition, priorities and control order. University of Illinois Engineering-Psychology Laboratory Technical Report EPL-81-4/ONR-81-4).

Wandmacher, J. (1981). Contour effects in figure perception. Psychological Research, 43, 347-360.

Watson, B. L. (1972). The effect of secondary tasks on pilot describing functions in a compensatory tracking task. University of Toronto Institute for Aerospace Studies Technical Note No. 178.

Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: An object superiority effect. Science, 186, 752-755.

Wewerwinke, P. (1976). Human monitoring and control behavior. Netherlands National Aerospace Laboratory Technical Report, NLR TR 77010 U.

Whitaker, L. (1979). Dual task interference as a function of cognitive processing load. Acta Psychologica, 43, 71-84.

Wickens, C. D. (1976). The effects of divided attention on information processing in tracking. Journal of Experimental Psychology: Human Perception and Performance, 1, 1-13.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and performance VIII. Hillsdale, N.J.: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman, & D. R. Davies (Eds.), Varieties of attention (pp. 63-102). New York etc.: Academic Press.

Wickens, C. D., Isreal, J., & Donchin, E. (1977). The use of the event related cortical potential as an index of task workload. Proceedings 19th Annual Meeting of the Human Factors Society, San Francisco, Calif..

Wickens, C. D., & Kessel, C. (1979). The effect of participatory mode and task workload on the detection of dynamic system failures. IEEE Transactions on Systems Man & Cybernetics, 13, 21-31.

Wickens, C. D., & Sandry, D. L. (1982). Task hemispheric integrity in dual task performance. Acta Psychologica, 52, 227-248.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility & resource competition between modalities of input, control processing and output: Testing a model of complex performance. Human Factors, 25, 227-248.

5.4.3. SPATIAL PROCESSING
by Hans-Willi Schroiff

The ability to deal with spatial relations has been traditionally regarded as an essential component of human cognitive functioning. Tests that supposedly tap this component have been incorporated in psychological tests of human intelligence. If we speak about the 'ability' to internally manipulate spatial relations as a fundamental part of the system that transforms and processes environmental information we assume that people differ reliably on this dimension.

According to Cooper & Regan (1985) spatial ability is defined as '... competence in encoding, transforming, generating, and remembering internal representations of objects in space and their relationsships to other objects and spatial positions'.

Tests of spatial aptitude also represent an interesting research area if one is interested in the attentional and perceptual correlates of human performance: The information processing demands of these tests have major communalities with basic perceptual processes and, unlike verbal materials spatial tasks do not depend that strongly on acquired specific knowledge.

In the following we first shall follow the development of the concept of 'spatial aptitude' through three successive psychological frameworks: Factor analysis, the information processing paradigm, and the so-called strategy approach.

SPATIAL APTITUDE AND CORRELATIONAL APPROACHES: correlating performance differences

It is not the aim to give a full account of the numerous studies within the correlational approach that have dealt with spatial aptitude. Instead we will try at least to give a sketchy outline of some major research programs. Factor analysis is concerned with relationships between individual differences in the performance of a large sample of tasks (see Fleishman & Quaintance, 1984). Factors that could be characterized as 'spatial' already appear in the early factor-analytic literature (e.g. Thurstone, 1938; McFarlane, 1925): 'Spatial visualization' was one of Thurstone's 'Primary Mental Abilities' (Thurstone, 1938). In the work of Cattell (1941, 1963) spatial factors were incorporated and referred to as determinants of the so-called 'crystallized' intelligence since a decline was observed with brain damage and aging. Guilford (1977) organized intellectual abilities in his classical factor-analytic 'structure of intellect' model along the three dimensions 'contents' (input), 'operations' (processing), and 'products' (output). Within this structure facets of spatial aptitude can be easily located. Pawlik (1973) defined a factor 'visual perception' that was supposed to reflect individual differences in tests involving visual stimulus material. The test scores loading on this factor were based on simple tests of perceptual speed as well as complex tests of spatial visualization and perceptual closure. French, Ekstrom, & Price (1963) included 'spatial scanning' in their 'Kit of Reference Tests for Cognitive Factors'. Based on Ekstrom's (1973) results Dunnette (1976) postulated 10 factors that included 'spatial orientation' and 'spatial visualization'. Harman (1975) expanded the work of French et.al. (1963) and identified 23 cognitive and temperamental factors with accompanying reference tests. Three factors (spatial orientation, spatial scanning, spatial visualization) are presumably related to spatial abilities.

Thus there is little doubt that within the factor-analytic reserach tradition spatial aptitude constitutes one of the central determinants of cognition. Spatial aptitude tests have been used as predictors of performance in both scholastic and industrial settings. The predictive validity of measures of spatial aptitude has been summarized by McGee (1979): Traditional spatial tests show substantial correlations with course grades in mechanical drawing, shop courses, art, mechanics, mathematics, and physics. In the area of performance in industry spatial tests have been predicting success in engineering, drafting, design, and other mechanically oriented areas.

More recently Lohman (1979) has reanalyzed most of the major U.S. factor analytic work on spatial aptitude. His results suggest a broadly defined spatial factor with several correlated subfactors. Three of these subfactors were consistently found in his reanalyses (following quotations from Lohman & Kyllonen, 1983, p.111):

- Spatial relations

  '... This factor is defined by test such as Cards, Flags, and Figures (Thurstone, 1938). These tests are all parallel forms of one another, and the factor only emerges if these or highly similar tests are included in the battery. Although mental rotation is the common element, this factor probably does not represent speed of mental rotation; rather, it represents the ability to solve such problems quickly, by whatever means'.

- Spatial orientation

  '... This factor appears to involve the ability to imagine how a stimulus array will appear from another perspective. In the true spatial orientation test, the subjects must imagine that they are reoriented in space, and then make some judgments about the situation. There is often a left-right discrimination in these tasks, but this discrimination must be made from an imagined perspective. However, the factor is difficult to measure since tests designed to tap it are often solved by mentally rotating the array rather than reorienting an imagined self'.

- Visualization

  '... The factor is represented by a wide variety of tests, such as paper folding, Form Board, Surface Development, Hidden Figures, Copying, and so forth.... The tests that load on this factor, in addition to their spatial-figural content, share two important features: they are all administered under relatively unspeeded conditions, and most are much more complex than corresponding tests that load on the more peripheral factors. Tests designed to measure this factor usually fall near the center of a two-dimensional scaling representation, and are often quite close to tests of Spearman's g or Cattell's Gf.

Lohman (1979) has extended his reanalyses by applying multidimensional scaling and cluster analysis to the set of factors and subfactors for spatial aptitude. It was found that performance in spatial aptitude tasks was related to the ability to encode, remember, transform, and discriminate spatial stimuli (see also Lohman & Kyllonen, 1983). Factors such as Closure Speed (i.e. speed of matching incomplete visual stimuli with their long-term memory representations), Perceptual Speed (speed of matching visual stimuli), Visual Memory (short-term memory for visual stimuli), and Kinesthetic Judgment (speed of making left-right discriminations) may represent individual differences in the speed or efficiency of some of these basic cognitive processes. Furthermore, the results imply a theoretical structure that can be described as the 'speed-power'- or 'simple-complex'-dimension. Assuming that spatial aptitude involves the selection and sequencing of elementary mental processes as matching, identification, transformation etc. Lohman (1979) showed that tests tend to be less speeded and more correlated with measures of reasoning the higher the complexity of the test items i.e. the more elementary information transformations have to be applied on the visual code. As item complexity increases the importance of speed is less and the tests load on more power -related factors, such as reasoning. According to Lohman & Kyllonen (1983) this implies that there may be different mental transformations in the various types of items that are supposedly tap spatial aptitudes. Individual differences in the speed of solving simple spatial problems may be largely independent of individual differences in the ability to correctly solve difficult spatial problems simply because qualitative and quantitative differences exist with regard to the number and sequence of elementary information processes.

There are a number of problems with the factor -analytic approach. Aside from general hesitations towards factor analysis as a tool to test hypotheses there is still a lot of disagreement on the number of subfactors needed to decribe spatial aptitude (see e.g. Lohman, 1979; McGee, 1979). The reasons for the inconsistencies are manifold: Different factoring methods lead to a different factor structures which in turn lead to different interpretations; even minor changes in task settings and in the choice of dependent variables may result in a change of the factor structure; the choice of tasks to be incorporated in the task sample affects the number and the nature of the resulting factors. Furthermore, the factor-analytic approach is based on two questionable assumptions (see Cooper & Mumaw, 1985):

(1) Factors stand for mental processes that are assumed to be common for the group of tests that load highly on a particular factor. The methodology, however, provides no way of testing this implicit hypothesis on the process of generating an outcome. The implicit use of untested rational process models render the factor-analytic approach somewhat arbitrary.

(2) It is tacitly assumed that solution strategies are invariant over the subjects. Different solution strategies -- i.e. a different selection and sequencing of elementary information processes -- however, lead to different factor structures. For example, the changes of factor structure with age could simply be explained by variations of solution strategies

that are tied to different age levels. Since factor analysis only evaluates the test score as the outcome of task performance and not the process, little is known about the variability of solution strategies and their effects on performance scores.

SPATIAL APTITUDE AND THE INFORMATION PROCESSING APPROACH: assessing quantitative interindividual differences

The information-processing approach in cognition is derived from the general model of information theory (see Lachman, Lachman & Butterfield, 1979). Thus Levine & Teichner (1973) defined a task as a transfer of information between an information source and a receiver in any system that can be construed as an information channel. Task performance is defined as a transfer of information between components; an operation on information or on data within a component is called a process. The analogy of the information-processing approach with a Turing machine is striking: human behavior is viewed as the instantiation of the symbol-manipulating capacity of a general purpose machine. Furthermore, cognitive processes are embedded in time, their duration is informative. Symbol manipulation is supposed to take place in processing stages some of which can be isolated by chronometric methods. The 'additive factor logic' is a striking example for a methodology that grew out of the information processing paradigm (Sternberg, 1969; Sanders, 1980). The chronometric methodologies imply, however, that people perform a task basically the same way. This is one of the reasons why the methodology is only applicable in a limited task domain (e.g. choice reaction tasks). While the factor-analytic approach only makes inferences about the processing requirements of a task by analyzing the correlation patterns with other equally unmodeled tasks the information processing approach represents the other extreme. The explicit assumption is that each subject's performance is best described by a number of elementary processes in the same sequence. The ultimate aim consists in identifying the subset of processes that explain individual differences in task performance. Usually this is accomplished by decomposing reaction times by means of the variation of task variables and to correlate these latency estimates with aptitude test scores. Variation across individuals can only be explained by variations in the speed or efficiency needed to perform inferred processing stages. Therefore Cooper & Mumaw (1985) refer to this approach as the '... identification of quantitative individual differences'.

The factor-analytic approach could be characterized by a lack of models of task performance; a major advantage of the information-processing approach is its need for building models which facilitate the identification of task components in terms of basic cognitive processes. The work of Roger Shepard and his colleagues represents a classical example of the information-processing approach in the investigation of spatial aptitude (see e.g. Cooper & Shepard, 1973; Shepard & Feng, 1972; Shepard & Metzler, 1971).

Shepard & Metzler (1971) required subjects to determine as rapidly as possible whether pairs of perspective two-dimensional drawings of three-dimensional objects had the same shape or were each other mirror images. Furthermore, the objects differed in angular disparity. The results showed that the time to make a same-different judgment was a linear function of the degree of angular disparity between the objects.

In this experiment and in numerous others that basically show the same results the underlying performance model assumed that, upon encoding the first object a mental rotation is performed on one of the stimuli in order to rotate its mental image to the same orientation as the other object and to compare the generated mental image to the actual representation of the second object. On the basis of the RT-distribution it was inferred that this mental rotation is performed analog to a physical rotation. A transformation on a mental image was postulated which assumed a structural isomorphism between the rotation of a physical and a mental stimulus -- between perception and imagery. The processing rate could be inferred from the slope of the RT-function (55. - 60./sec) which was conceived as highly dependent on the degree of mental rotation. The intercept of the function relating RT and angular disparity was conceived as an estimate for the duration of the processes independent of spatial manipulation.

From the Shepard & Metzler (1971) study one might gain the impression that the mental rotation is indeed responsible for the observed individual differences since overall RT showed a strong dependency on angular disparity. However, performance variation can be localized anywhere within the chain of processing stages.

A study by Egan (1978) investigated the relationship between accuracy and latency measures of spatial aptitude. Correlational analyses clearly demonstrated that accuracy scores were highly interrelated and that latency measures were highly interrelated, but accuracy and latency did not seem to measure the same aspect of behavior. A subsequent factor analysis revealed two distinct factors with one loading on accuracy and the other on speed.

Mumaw, Pellegrino, & Glaser (1980) required their subjects to rapidly determine whether an array of pieces on the right can be used to assemble a completed puzzle on the left. They developed an information-processing model for this task that represented a piece-by-piece processing loop until a mismatch is detected or until all pieces are checked. Five item types were created that differed with respect to (a) the number of pieces (2-6) and (b) the arrangement of the array on the right (scrambled and rotated, scrambled, rotated, separated, holistic). RT-functions reflect the effects of both experimental variables. The pattern of results (see Mumaw et.al., 1980) implies that there are two independent sources contributing to high ability in this particular task. One is reflected by the speed of search, speed of encoding and comparison, the other is the ability to rotate pieces accurately.

According to Cooper & Mumaw (1985) these results suggest the majority of traditional measures of spatial aptitude is not closely related to the speed of mentally transforming an internally generated representation but has to do more with the speed and the quality of the encoding and comparison processes -- especially when the items are more complex: as the task becomes more complex and more transformations are applied to a single representation, the quality and stability of that representation should become more important (see also the argument of Lohman, 1979).

Elsewhere we have argued (Schroiff, 1983; Schroiff, in press) that process models require specific process methodologies to be tested. Under certain circumstances analysis of eye-movements provides a useful tool to observe the time characteristics of task performance dealing with visually presented stimulus materials. Just & Carpenter (1978) collected eye-movement parameters (i.e. fixation sequences and fixation durations) of subjects who were solving items of the Shepard & Metzler type. Their information processing model for the Shepard & Metzler task involved basically three consecutive processing stages: The 'search' stage concerned with the selection of a stimulus segment that is to be transformed, the 'transform and compare' stage involves stepwise mental transforming of the selected stimulus segment and a comparison with the reference item that is supposed to remain unrotated. In the final 'confirmation' stage it is decided by further cross-checks whether other segments can also be brought into congruence by the rotation process. In order to obtain latency estimates for these three stages the processing operations were tied to observable eye-movement behavior: 'Search' is defined as the time that elapses prior to the repeated switching between identical stimulus segments which defines the 'transform and compare' stage. 'Confirmation' is indexed by fixations of segments that are not fixated during the 'transform and compare' stage. In the figure below the mean stage latencies for 'same' trials are plotted as a function of angular disparity:

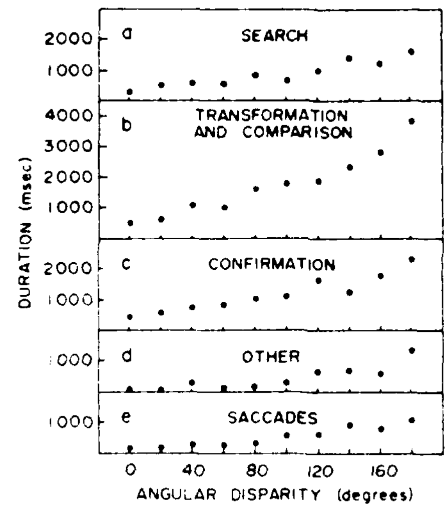


Figure 1: Mean duration of vaious pprocessing stages in Same trials as a function of angular disparity, with complex three-dimensional stimuli. (From Just & Carpenter, 1976.)

The figure shows that the latency to make a 'same'-'different' judgment in the Shepard & Metzler task is at least composed of the three above mentioned processes. Thus the individual slope of the RT-function does not only reflect differences in mental rotation but also differences in search for a segment that can be rotated. Again, however, this seems to depend on the complexity of the stimulus items. In a subsequent study (Carpenter & Just, 1978) the question was raised whether stimulus complexity exerts an influence on the processing stages assumed for mental rotation tasks. Instead of perspective line drawings two-dimensional dot patterns were used as stimuli. Otherwise the experimental conditions remained identical. Once more the results showed a linear increase of total RT with increasing angular disparity. However, in this case 'search' and 'confirmation' did not explain any variance of the total RT: Angular disparity only showed an influence on the 'transformation and comparison' process. According to Carpenter & Just (1978, p. 123) '.... simpler figures did not cause confusion between the segments and hence there was no increase in initial search duration' and '.... because the figures were strictly treated as two-dimensional figures and because the main segments of each figure were discriminable, the confirmation process was unnecessary'.

Again we can make a case that the complexity of the stimulus materials exerts an influence on the information processing structures and functions involved and, as an important consequence, has an impact on the validity of tests based on these stimuli with regard to the prediction of spatial aptitude.

In summary, the information-processing approach to spatial aptitude has provided a number of interesting insights into the microstructure of mental rotation phenomena. However, it is assumed that all subjects carry out the same mental operations to perform a mental rotation task in the same sequence and that differences in spatial aptitude are only reflected in speed or efficiency differences by which these operations can be carried out. In this respect it is less important whether the encoding operation or the transformation of the mental image explains the major portion of reaction time. Each of these explanations would suggest that individual differences are attributable to some inherent, yet ill-defined trait in the perceptual-memory system like the richness or stability of some mental representation. It remains unclear whether the quality of this image is based on specific encoding, on speed of search or on whatever other factor.

There can be no doubt that differences in the quality of the mental representation exist and that these differences explain a major portion of variance in most mental rotation paradigms. As, however, the degree of task complexity or difficulty increases other cognitive functions come into play which have some communalities with those employed in reasoning tasks. In these cases the 'perceptual portion' of spatial aptitude is dominated by the operations of a mental executive that is responsible for the appropriate selection and the appropriate sequencing of basic information processes. This is clearly demonstrated by the function of the search process in the Shepard & Metzler task where prior to rotation a decision has to be made which segment is the most promising for being rotated.

Thus it may well be that performance in mental rotation is determined to a major extent by a repertoire of strategies that subjects employ to solve problems. This issue will be addressed in the following section.

SPATIAL APTITUDE AND STRATEGY SELECTION: Qualitative individual differences

Information-processing frameworks like the additive factor logic play an important role as long as the the task leaves only limited strategic freedom. By appropriate choice of tasks and experimental control reaction time differences are reliable performance differences for one consistent behavioral mode. However, even in most laboratory tasks this seems hard to accomplish (see Debus & Schroiff, 1984). For instance, in mental arithmetic, different solution strategies may lead to different performance scores which entail different factorial structures on the one hand or — depending on the paradigm -- different inferences based on reaction times on the other hand. Neither the factor-analytic approach nor the information

-processing framework has seriously considered the possibility that people may perform in qualitatively different ways in the sense that the same performance result -- e.g. a 'same'-'different' judgement -- can be obtained by means of radically different procedures. Schroiff, Borg, & Staufenbiel (in press) provided a striking example. In a paired-comparison task their subjects made similarity judgments about two simultaneously presented rectangles while their eye-movements were recorded. Based on eye-movement parameters subjects could be classified as either 'holistic' (i.e. using the integral dimension 'area' as a basis for comparison) or 'analytic' (i.e. using the separable dimensions 'width' and 'height'). Yet the multidimensional scaling configurations of 'analytic' and 'holistic' subjects were virtually indistinguishable. It appeared that reaction times and eye-movement parameters are indicative of process characteristics of similarity judgements but are no predictors of the result of the judgmental process. Debus & Schroiff (1985) demonstrated different solution strategies in a digit-symbol-substitution task that relied on either the build-up and use of an internal store or on the rapid access to an external store. Depending on the strategy employed the test has a different predictive validity.

In summary, it could be that differences in spatial aptitude may be simply related to either differences in global strategy or flexibility in strategy selection which entails differences in a repertoire of strategies and decision strategies for the selection of the most effective strategy for the task at hand. This does not mean, however, that the importance of the speed of the underlying processing operations is denied but the variance in higher mental processes ('mental executive') may play a more important role than usually assumed. The question remains why strategies as major determinants of behavior have been so much neglected. Lohman & Kyllonen (1983) give a tentative answer: '....The research community has occasionally acknowledged the problem of alternative solution strategies, but never taken the possibility too seriously, since it would necessitate a serious rethinking of the meaning of test scores and, more generally, of all experimental tasks.'

This is especially surprising when we consider what impact different solution strategies have for the two framworks that we already considered: Individual differences in solution strategy are a basic challenge for factor analysis. The most likely outcome for a task sample that allows for different solution strategies is an overestimation of the factorial complexity of the test. In that case it cannot be decided whether this factorial complexity is caused by between-variance or within-variance of strategies. French (1965) demonstrated that different strategies which could be labeled either 'analytic' or 'global' as assessed in a posteriori interviews yielded different factor loadings in some psychometric tests. The information-processing paradigm is also challenged to the extent it has the basic assumption that the task is performed in the same way by all subjects. Here again the examples given by Debus & Schroiff (1984) speak for themselves.

Let us consider some examples. The first concerns an experiment of Putz-Osterloh (1977) which is discussed despite the fact that it cannot be easily related to one of the proposed frameworks. Starting from results of the correlational approach her research focused on a problem solving perspective of spatial aptitude employing a 'cubes comparison' task (Amthauer, 1953).

Although the analyses were not based on an explicit process model, performance could be predicted on the basis of a theoretical task analysis. Putz-Osterloh (1977) identified three possible ways of solving the 'cube comparisons': (1) area comparisons: the same-different judgment can be made by simply comparing the three visible sides (2) area comparisons +: a same-different judgment can be made by comparing: visible sides and one relation between these sides (3) spatial comparison: a same-different can be made by checking the identity of two visible sides and imagining an new third side. It becomes clear from this analysis that the stimulus material in the 'cubes comparison' task is far from being homogeneous and thus simply may not measure a specific ability. In fact, the performance data (reaction times, error proportions, and eye-movement parameters) clearly indicate two separate classes of allegedly spatial test items, where one class (area comparisons) does not require spatial transformations at all. These tasks can indeed be performed by simple area comparisons while spatial operations only seem to be necessary for the item category 'spatial comparisons'. Furthermore, this study demonstrated that subjects may react to changes in stimulus requirements by choosing a different and more efficient strategy. Subjects employing a 'feature-analytic' strategy had some difficulties in switching to the spatial strategy when this was required.

In various so-called spatial tasks subjects may employ one of two broad classes of strategies labeled 'holistic' and 'analytic' (Cooper, 1976, 1980, 1982; see also Cooper & Podgorny, 1976). Schroiff (1983) summarized a number of studies that all demonstrate reliable differences between holistic and analytic subjects. This could be demonstrated for rather simple tasks (e.g. Cooper, 1976) as well as for more complex tasks like the 'Advanced Progressive Matrices' (e.g. Hunt, 1974). It appears also that subjects are flexible in applying these strategies. Cooper (1980, 1982) reported that 'holistic' subjects could switch to the 'analytic' mode when a task demands required that particular strategy. The reverse, however, seems to be less likely so that persons with a more 'holistic' mode seem to be more flexible.

Just & Carpenter (1983) have summarized possible strategies in mental rotation tasks in terms of their theory how people solve problems on psychometric tests of spatial ability. They assume that spatial information is coded with respect to a cognitive coordinate system. In order to explain individual differences they suggest that the use of different coordinate systems may explain individual differences in spatial ability, as well as strategic differences in spatial tasks. They suggested the following strategies:

- mental rotation around standard axes
  This form of mental rotation is most frequently discussed in the psychological literature (e.g., Cooper & Shepard, 1973). The axis of rotation is one of the usual three axes of space, as defined by the visual environment, gravity, picture plane etc. These frames of reference are outside the object that is being rotated.

- mental rotation around task defined axes

  This form invoves mental rotation around an arbitrary axis that is particular useful for the task at hand. The process by which subjects determine this axis of rotation becomes interesting if the axis is determined by the problem.

- comparison of orientation-free descriptions

  Subjects using this strategy code the relation of two elements on the left cube in a Cubes Comparison task and then determine whether this relation can also be found on the right cube. In this case no mental rotation is involved (orientation-free).

- perspective change

  The use of this strategy entails mentally changing the representation of the observer's position relative to the object and hence his or her view of the object, but keeping the representation of the object's orientation in space constant (see 'Schlauchfiguren'). The axis-finding process becomes a decision of which view to take of the object.

Based on their process model Just & Carpenter (1978) analyzed how people perform in the Cubes Comparison task. The final aim was to determine which processes distinguish subjects of high spatial ability from subjects with low spatial ability. Again a process methodology (eye-movement recording) was used to trace the sequence and duration of the component processes.

Reaction time data showed longer reaction times for subjects with low spatial ability (low spatial subjects). Groups of high spatial and low spatial subjects reported both rotation strategies and the strategy of orientation-free descriptions. For the latter strategy the pattern of reaction times for the postulated steps differed. It was found that the two subject groups differed with respect to 'initial rotation' (low-spatial subjects take longer) and 'confirmation' (low-spatial subjects take longer). No differences between the groups were observed with respect to the search process. The difference in the rotation strategies employed by the two performance groups can be viewed in terms of a difference in the cognitive coordinate system: Low-spatial subjects almost never used a cognitive coordinate system that did not closely correspond to the cubes' axes or to the axes of the visual environment. In addition high-spatial subjects seem to have a faster rotation rate. The reasons remain unclear. It may be possible that a faster rotation rate is caused by (1) faster execution of a basic mental operation (2) a more economical code to represent the figure (3) a larger rotation angle per step.

One single high-spatial subject who employed an orientation-free description strategy showed response times that were considerably slower than the average of the high spatial subject but still slightly faster than the low spatial subjects. Just & Carpenter (1985) pinpoint the problem: '... The existence of this strategy illustrates that tasks ostensibly requiring spatial manipulation can be effectively performed without manipulation if the appropriate cognitive coordinate system is used.' This means that no reliable inferences about the processing mode can be made on the basis of performance measures like RT or error proportion.

The 'perspective change' strategy could not be observed among Just & Carpenter's subjects although theoretically this strategy can be used to solve Cubes Comparison items. In this strategy the object's orientation in space is kept constant, but there is a change in the representation of the viewing point.

It should be clear from the results of Just & Carpenter (1985) and Putz-Osterloh (1977) that the items in most tests of spatial aptitude allow for more than one solution strategy — especially when the item pool is heterogeneous. It may even be the case that different items or different versions of the test invoke a specific strategy (Just & Carpenter, 1986, see also Barratt, 1953) leading to both within-subject and between-subject strategy variation. All this requires more detailed analyses of inter- and intraindividual differences in solution strategies based on a theory that tries to explain strategy choice and process characteristics of the individual strategy.

Strategy choice seems to depend to a major degree on the characteristics of the individual test item — especially on its difficulty or complexity that for the moment we put on one level with the degree of strategical freedom and the probability of errors. In discussing the results of Lohman (1979) we have already pointed out that more complex tasks require more complex information processing so that the repeatedly observed correlations with reasoning tests are not surprising (see Steller & Stürmer, 1984).

OUTLOOK

We may conclude on the basis of this review that valid assessment of spatial aptitude is a complicated affair for two reasons:

(1) HETEROGENEOUS ITEM PROBLEM

We may conclude that different psychological functions come into play when the level of task complexitxy in spatial tasks changes: The more complex the task the more likely reasoning factors come into play. In addition the probability increases that the subjects employ more than one solution strategy leading to the

(2) HETEROGENEOUS STRATEGY PROBLEM

There are basically two ways out of this dilemma:

HOMOGENEOUS ITEMS/HOMOGENEOUS SOLUTION STRATEGY

(1) design tasks where solution strategies which are not based on spatial manipulation do not lead to successful task performance (increase item homogeneity and strategy homogeneity).

Thus Putz-Osterloh (1977) showed that for a subset of cube comparisons items the strategy of successive feature comparison did not lead to a correct solution. Gittler (1984) has proposed the application of the Rasch model

HETEROGENEOUS ITEMS/HETEROGENEOUS SOLUTION STRATEGIES

(2) design tasks based on empirical evidence how various solution strategies manifest themselves in the pattern of results. In the test situation the solution strategy can be inferred from the pattern of results and thus become part of the psychometric test. This is particularly interesting for those situations where the strategical control of behavior has to be assessed.

Both problems are related to each other in the sense that solution strategy is a function of the interaction between item characteristics and person characteristics. We believe that it is useless to sort people into typological categories because this is equivalent to assuming that the strategy once chosen is is a consistent feature of the person. However, we have tried to argue that persons can be characterized by their flexibility to employ strategies dependent on item characteristics. On the other hand this implies that we cannot sort items into the same category because this would mean that all persons solve the items in the same way which is equally improbable within the proposed framework. What is required are models for the within-shifts in solution strategy.

Methodological requirements

The identification of strategies, however, entails a number of methodological problems.

(1) If separate strategies are postulated a priori there should be a separate process model for each strategy. Here again the work of Debus & Schroiff (1984) is a good example.

(2) Process models require process methodologies to be tested. It has been shown that eye-movement analysis is a promising reserach-tool for two reasons: First, if the subject is allowed to perform the task in the usual way eye-movement analysis may help in combination with a process model (see Schroiff, in press) to identify the various strategies in spatial tasks. In this case strategy is an independent variable whose influence is estimated ex-post-facto. In this quasi-experimental approach the notion 'strategy' may be used to describe different action patterns found in the data. Second, in an experimental approach eye-movement analysis may serve as an experimental control when strategy becomes an a priori defined independent variable. If e.g. the subject is instructed to follow a particular strategy eye-movement monitoring allows for a direct experimental control whether the subject behaved according to the instructions.

It should also be clear that eye-movements will have to be cross-validated by employing other methodologies like performance data and thinking aloud protocols in order to facilitate the interpretation of complex eye-movement patterns.

High- and low-spatial aptitude groups may not only be characterized by their processing strategies (tactical aspect), but also in the richness of their spatial representations and in their ability to maintain a complex spatial structure in memory (ability aspect, which in turn may be related to the strategy employed). Paradigms should be developed that explicitly test this essential requirement for spatial aptitude. Individuals high in spatial aptitude may have a diverse set of available strategies and be efficient and flexible in strategy application.

The challenge for future research is to design experiments where the solution strategy of the subject becomes visible for each item. Only in that case the investigator can evaluate the results and the generalizability of the processing models. As Cooper & Mumaw (1985) have pointed out additional methodologies are needed to separate out different strategies. A further possibility consists in the construction of test materials that invoke a different strategy and thus make strategic differences an additional diagnostic tool.

It would seem rewarding to identify stimulus characteristics that govern the choice of solution strategy and apply this knowledge in the construction of tests that systematically vary these characteristics. In that case different strategies would not decrease but increase the validity of a test.

REFERENCES

Cattell, R.B. (1941). Some theoreticl issues in adult intelligence testing. Psychological Bulletin, 38, 592 (abstract).

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54, 1-22.

Cooper, L.A. (1976). Individual differences in visual comparison processes. Perception & Psychophysics, 19, 433-444.

Cooper, L.A. (1980). Spatial information processing: Strategies for research. In: R. Snow, P.A. Federico, & W.E. Montague (Eds.). Aptitude, learning, and instruction: Cognitive process analyses. Hillsdale, NJ: Erlbaum.

Cooper, L.A. (1982). Strategies for visual comparison and representation: Individual differences. In: R.J. Sternberg (Ed.). Advances in the Psychology of Human Intelligence (Vol.1). Hillsdale, NJ: Erlbaum.

Cooper, L.A. & Podgorny, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity: Journal of Experimental Psychology: Human Perception and Performance, 2, 503-514.

Cooper, L.A. & Shepard, L.N. (1973). Chronometric studies of the rotation of mental images. In: W.G. Chase (Ed.). Visual information processing. New York: Academic press.

Dunnette, M.D. (1976). Aptitudes, abilities, and skills. In: M.D. Dunnette (Ed.). Handbook of Industrial and Organizational Psychology. Chicago: Rand McNally.

Egan, D.E. (1978). Characterizing spatial ability: Different mental processes reflected in accuracy and latency scores. (Research Rep. No. 1224). Pensacola, Florida: Naval Aerospace Medical Research Laboratory.

Ekstrom, R.B. (1973). Cognitive factors: Some recent literature (Tech.Rep. No.2, ONR Contract N00014-71-C-0117, NR 150-329). Princeton, NJ: Educational Testing Service.

Fleishman, E.A. & Quaintance, M.K. (1984). Taxonomies of Human Performance. The Description of Human Tasks. New York: Academic Press.

French, J.W. (1965). The relationship of problem-solving styles to the factor composition of tests. Educational anu Psychological Measurement, 15, 9-28.

French, J.W., Ekstrom, R.B., & Price, L.A. (1963). Manual for kit of reference tests for cognitive factors. Princeton, NJ: Educational Testing Service.

Guilford, J.P. (1977). Way beyond the IQ. Buffalo, NY: Creative Education Foundation.

Harman, H.H. (1975). Final report of research on assessing human abilities (PR-75-20). Princeton, NJ: Educational Testing Service.

Hunt, E. (1974). Quote the Raven? Nevermore! In: L. Gregg (Ed.) Knowledge and Cognition. Hillsdale, NJ: Erlbaum.

Just, M.A., & Carpenter, P.A. (1976). Eye fixations and cognitive processes. Cognitive Psychology, 8, 441-480.

Just, M.A. & Carpenter, P.A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. Psychological Review, 92, 137-172.

Levine, J.M., & Teichner, W.H. (1973). Development of a taxonomy of human performance: An information-theoretic approach. JSAS Catalog of Selected Documents in Psychology, 3, 28 (Ms. No. 325).

Lohman, D.F. (1979). Spatial ability: A review and reanalysis of the correlational literature (Tech. Rep. No. 8). Aptitude Research Project; School of Education, Stanford University.

McFarlane, M.A. (1925) A study of practical ability. British Journal of Psychology Monograph Supplement, Whole No. 8.

McGee, M.G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychological Bulletin, 86, 889-918.

Mumaw, R.J., Pellegrino, J.W., & Glaser, R. (1980). Some puzzling aspects of spatial ability. Paper presented at the annual meeting of the Psychonomic Society, St. Louis.

Pawlik, K. (1973) Faktorenanalytische Persönlichkeitsforschung. In: Kindlers Enzyklopädie der Psychologie des 20. Jahrhunderts. Zürich.

Putz-Osterloh, W. (1977). Über Problemlöseprozesse bei dem Test Würfelaufgaben aus dem Intelligenzstrukturtest IST und IST-70 von AMTHAUER. Diagnostica, 23, 252-265.

Schroiff, H.W. (in press). Zum Stellenwert von Blickbewegungsdaten bei der Analyse kognitiver Prozesse. To appear in: Zeitschrift für Psychologie.

Schroiff, H.-W. (1983). Experimentelle Untersuchungen zur Reliabilität und Validität von Blickbewegungsdaten (Experimental studies on the reliability and validity of eye-movement data). Unpublished Doctoral Dissertation. Rheinisch-Westfälische Technische Hochschule Aachen, Federal Republic of Germany.

Schroiff, H.-W., Borg, I., & Staufenbiel, T. (in press). Similarity judgements, distances, reaction times, and eye-movements in the perception of rectangles. To appear in: Archiv für Psychologie.

Shepard, R.N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. Science, 171, 701-703.

Steller, M. & Stürmer, M. (1984). Space, reasoning oder was? Zur Validität des 'Raumvorstellungstests' LPS 8/PSB 7. Zeitschrift für experimentelle und angewandte Psychologie, 31, 671-692.

Thurstone, L.L.(1938). Primary mental abilities. Psychometric Monographs,1.

Zimmermann, W.S. (1954). The influence of item complexity upon the factor composition of a spatial visualization test. Educational and Psychological Measurement, 14, 106-119.

5.4.4. PERCEPTUAL MOTOR SPEED AND CHOICE REACTION PROCESSES
by Jan Theeuwes

INTRODUCTION.

The idea underlying reaction time measurement is that mental processes are embedded in real time. This implies that it is possible to relate mental events to physical measures. This approach can only be useful if reaction time can be decomposed into a finite number of functional subunits or stages which, unobservable by themselves, can be inferred through manipulation of tasks or task variables. A main aim of studying RT is concerned with this stage analysis of reaction processes (see Sternberg, 1969; Sanders, 1980a). The approach leads to construction of a sequence of individual stages, the combination of which results in a model which describes the stage structure of the reaction process. This step should be followed by an analysis of processes within stages.

This paper is concerned with a concise outline of stage analysis of choice reaction processes along the lines of different theoretical notions. At least four different stages appear to be involved in choice reactions: first, reception of the signal by a sense organ and conveyance of the data through the afferent nerves to the brain; second, identification of the signal; third, choice of the corresponding response; and fourth, initiation of an action that constitutes the response (Welford, 1980).

The first section is concerned with stage analysis along the lines of the additive factor method (Sternberg, 1969). An attempt is made to relate the processing durations of the individual stages to psychological meaningful concepts. The additive factor methodology of decomposing reaction times is an important topic in the current literature. It might not only provide a tool for distinguishing structural or "computational" mechanisms of information processing but one for analysing energetical resources as well (Sanders, 1981, 1983; Frowein, 1981a, 1981b; Gopher & Sanders, 1984). Yet from a theoretical point of view it has been questioned whether interpretations of the reaction process as inferred by the additive factor method are valid (Taylor, 1976; Stanovich & Pachella, 1977; Rabbitt, 1979; Hockey, 1979; Pachella, 1974).

Given the stage structure, as outlined in the first section, the second section will be devoted to the discussion of serial or parallel processing of the information flow which requires consideration of the distinction between automatic versus controlled processing.

THE ADDITIVE FACTOR METHOD (AFM).

Although in various applied situations it can be useful to measure reaction time without any theoretical background, this kind of approach is of little significance to information processing research. In order to be relevant to basic research it is neccessary to design an experiment in such a way that conclusions can be inferred about the relation between obtained variations in reaction time and variations in the durations of particular types of processing. What is needed are converging notions underlying the decomposition of the reaction time. Two common approaches are the subtraction method (Donders, 1868; see also Pachella, 1974) and the additive factor method (Sternberg, 1969). The subtraction method has been widely critized as an inadequate tool for stage analysis (Pachella, 1974; Sanders, 1980a). Its basic idea is that the duration of processing within a stage can be estimated when this stage is deleted. If one task consists of n stages and an other one of n-1 stages, the duration of the deleted stage can be inferred by subtracting the reaction times obtained at the two tasks. To apply this method one should obviously have prior knowledge about the sequence of events between stimulus and response. Hence, it requires a priori postulates of stages instead of inferring stages from reaction times. Another reason for criticism is the assumption of pure insertion, suggesting that the processing sequence of the stages is not affected when another stage is inserted. This is a matter of comparability of two different tasks. It is more plausible to assume that an insertion may change the whole processing structure.

The additive factor method is a more basic tool for "discovering processing stages" (Sternberg, 1969). The main distinction between this method and the subtraction method is that stages are actually inferred from the experimental data. This section will mainly deal with the AFM (for a detailed discussion of methodological issues see Section 3.2. in this report).

The AFM involves the following conceptions. First, it is assumed that the reaction time interval is filled with a sequence of independent processing durations, each of which represents a processing stage. Each stage performs a constant informational transformation; the output of this transformation is the input for the next stage. Second, the transformation produced by a stage is independent of processing durations of the preceeding stages. In addition, within a stage the time it takes to transform an input to an output (processing duration) is not related to the quality of that output. Thus, the quality of the input and output of each stage is independent of the stage in question and of those of the preceeding stages. The AFM is merely concerned with the processing durations of these stages and the factors affecting these durations.

Given these assumptions about stages the relationship between processing durations and experimental manipulations can be considered. If two experimental manipulations affect two different stages their effects on the reaction times will add. In a statistical sense this means that there are only main effects. The rationale for finding additive factors is that the effect of one variable does not appear to depend on the state of the other (Sanders, 1980a). Alternatively, if two experimental manipulations mutually modify each others' effects, the variables are likely to affect at least one commom processing stage. In a statistical sense this means that the effects of the variables interact: the effect of one variable is dependent on the state of the other.

Before summarizing the experimental results two methodological points should briefly be considered. First, it is important to take care that the experimental manipulation does not influence the structure of the task. Each stage has to produce an equal output across levels of experimental variables (Sanders, 1980a). For example, if a task becomes more difficult this should only affect the processing durations and not the quality of the output of each stage. Hence an experimental variable which redefines the experimental task can not be considered in terms of the AFM. This limit is analogous to the assumption of pure insertion as discussed in relation to the subtraction method. It implies that the AFM cannot be used as a method of analysis for all experimental manipulation of reaction time. An example of a clear violation of the assumption of equal output can be found in the experiment of Stanovich and Pachella (1977, experiment 1). Their extremely large effect of contrast variation (200 msec) suggests that the sensory stage probably produced distorted outputs. Second, it is important to note that the AFM can only be applied to stages and does not consider processes within stages. As Sternberg (1969) stated: "the additive factor method cannot distinguish processes but only processing stages" (p.369). For a detailed discussion of the methodological issues concerning the AFM see Sanders (1980a).

PROCESSING STAGES

In this section some experimental results concerning the stage analysis of choice reaction processes will be considered. The stages and the task variables are briefly discussed. Frowein (1981a) has presented a detailed model of the processing stages that together can account for reaction times in traditional choice reactions.

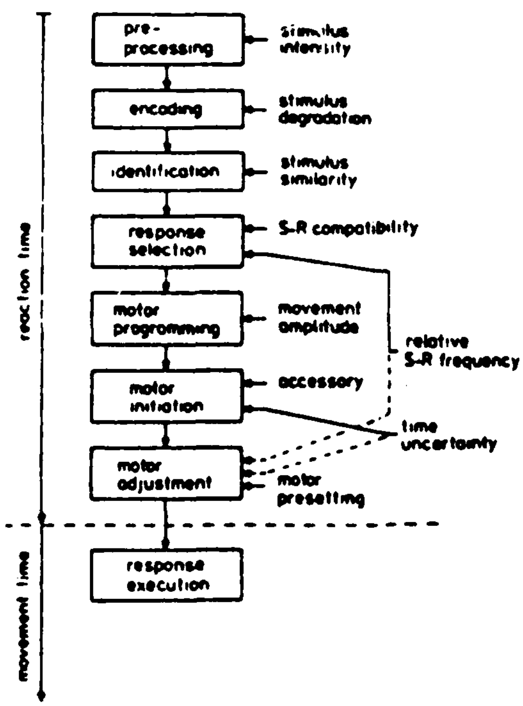


Figure 1: Task variables and inferred stages in the reaction process. (from Frowein, 1981a)

Table 1a gives a summary of some observed additive effects of taskvariables on choice reaction time, Table 1b shows the interactive effects between variables.

Table 1a: Summary of additive effects of task variables on visual choice reaction time.

TASK VARIABLES	AUTHORS
stimulus intensity + stimulus degradation	- Sanders (1980b) - Frowein (1981a)
stimulus intensity + stimulus similarity	- Pachella & Fisher (1969) - Shwartz et al. (1977)
stimulus degradation + stimulus similarity	- Shwartz et al. (1977)
stimulus intensity + S-R compatibility	- Sanders (1977) - Shartz et al. (1977)
stimulus intensity + time uncertainty	- Raab et al. (1961) - Sanders (1977)
stimulus intensity + rel. S-R frequency	- Stanovich & Pachella (1977, expt. 2 and 3)
stimulus degradation + S-R compatibility	- Frowein (1981a) - Sternberg (1969) - Shartz et al. (1977) - Sanders (1980b)
stimulus degradation + time uncertainty	- Frowein (1981a) - Wertheim (1979)
stimulus degradation + muscle tension	- Sanders (1980b)
stimulus similarity + S-R compatibility	- Pachella & Fisher (1969) - Shartz et al. (1977)
S-R compatibility + time uncertainty	- Frowein (1981a) - Posner et al. (1973) - Sanders (1977) - Spijkers & Walter (1985)

Table 1a continued

S-R compatibility + response specificity	- Sanders (1970)
S-R compatibility + muscle tension	- Sanders (1980b)
S-R compatibility + response duration	- Spijkers & Walter (1985)
rel. S-R frequency + time uncertainty	- Holender & Bertelson (1975)
time uncertainty + accessory	- Sanders (1980a)
time uncertainty + movement amplitude	- Frowein (1981a)
time uncertainty + response duration	- Spijkers & Walter (1985)
accessory + muscle tension	- Sanders (1980a)

Table 1b: Summary of interactive effects of taskvariables on visual choice reaction time

TASK VARIABLES	AUTHORS
S-R compatibility x rel. S-R frequency	- Frowein (1981a) - Fitts et al. (1963) - Broadbent and Gregory (1965) - Sanders (1970) - Theios (1975)
S-R compatibility x stimulus intensity	- Stanovich & Pachella (1977, expt. 1)
rel. S-R frequeny x stimulus intensity	- Miller & Pachella (1973) - Stanovich & Pachella (1977)
rel. S-R frequency x time uncertainty	- Bertelson & Barzeele (1965)
rel. S-R frequency x muscle tension	- Sanders (1980b)
rel. S-R frequency x response specificity	- Sanders (1970)
time uncertainty x muscle tension	- Sanders (1980b)
time uncertainty x accessory	- Frowein (1981a)
time uncertainty x S-R frequency x muscle tension	- Sanders (1980b)

PERCEPTUAL STAGES

The task variable "stimulus intensity" (contrast) is related to the luminance of the visual stimulus. "Stimulus degradation" is usually obtained by superimposing a checkerboard pattern (e.g. Sternberg, 1969). "Stimulus similarity" refers to the similarity between alternative stimuli. For example Shwartz et al. (1977) varied the slope of the upright lines in the capital letters A and H. The three "perceptual" variables appear to have additive effects on choice reaction time so it can be concluded that at least three perceptual stages are involved. There can only be some speculation regarding the nature of these stages. Preprocessing may represent some peripheral transport of sensory input, during the encoding stage a general feature analysis may occur and a final selection among possible stimuli alternatives may take place in the identification stage. It should be noticed that in particular the identification stage is based on fairly weak evidence and certainly deserves futher experimentation.

RESPONSE SELECTION STAGE

The response selection stage is influenced by S-R compatibility (spatial or semantic) and by relative S-R frequency. This last variable refers to the relative frequency of occurence of S-R pairs. If for example one pair occurs in 55% of the trials this results in a short reaction time for this pair. Figure 1 shows that relative S-R frequency interacts with S-R compatibility as well as with variables controlling motor presetting. It is relevant to add that additive effects between SR compatibility and signal degradation are well established in a number of studies.

MOTOR PROCESSING STAGE

Response execution variables are related to motor programming. The evidence regarding this stage is not yet well established, although it seems that movement amplitude has an influence on the reaction time (Fitts & Peterson, 1964). The idea underlying this stage is that ballistic movements (shorter than 220-290 msec) are programmed prior to initiating the response.

The task variables "accessory", "time uncertainty", "relative SR frequency" and "motor presetting" are thought to affect the motor stages "initiation" and "adjustment". The variable accessory refers to an irrelevant auditory stimulus which is presented simultanuously with a visual reaction stimulus. Although this auditory signal does not provide any further information, the reaction time is shorter when the accessory is present. Time uncertainty is related to the degree of uncertainty about the moment of presentation of the reaction signal. Manipulation of foreperiod duration (FPD) is a way to vary this uncertainty. Motor presetting refers to presetting of motor response prior to the reaction stimulus. A well known example of variation of presetting is instructed muscle tension (Sanders, 1980a). The figure shows the different interactive and additive relationships among these variables. With regard to the two motor preparation stages it is thought that the motor initiation stage reflects the subject's readiness to respond and that the motor adjustment stage constitutes the first part of response execution. (f.e. some muscular processes). Besides the additive and non-additive relations in choice reaction time there is some physiological evidence concerning CNV recordings to support the existence of motor preparation stages (Gaillard, 1978, 1980).

In the Frowein`s stage model (1981a), as outlined in figure 1, it is claimed that seven independent stages are involved in the choice reaction process. Orginally Sternberg postulated four stages: stimulus encoding, information processing and evaluation, response decision and response selection and evocation. Sanders (1980a) claimed that six stages are involved, whereby no distinction is made between the two motor preparation stages. It is apparent that an inflation of stages reduces the strength of the AFM. It seems that new stages can be only "discovered" if one can find stable relations between task variables. If a further fractionation of the reaction process will occur the stages are no longer psychological meaningful; what will be left is a one to one relation between a variable and a stage.

AN ELABORATON OF THE LINEAR STAGE MODEL

Regarding the nature of the reaction process the AFM assumes that reaction processes are one dimensional. This implies that the output of a stage can only serve as an input for one next stage. In this sense parallel processing of information can not be considered by the AFM, at least not between stages. Futhermore, as discussed earlier, the task variables are not allowed to influence the structure of the task. The linear stage model maintains that there is a fixed structure of computational stages, each cperforming an informational transformation. Given these very strict assumptions the linear stage model has been considered as a fully data driven model (Rabbitt, 1979; Hockey, 1979). In such model input starts up the sequence of stages, and processing takes place without any active influence from a central executive. In turn this would mean that cognitive states like motivation could not be included in the model. A model of information processing which cannot account for influences of cognitive states is so limited that it is fair to question the relevance of the model. Yet, the data driven nature of the linear stage model may not be fully correct. A clear example of an active influence on the reaction process is the effect of motor presetting (e.g. muscle tension) when the moment of the reaction signal can be predicted. Again the effects of relative signal frequency suggest active presetting prior to the arrival of the signal.

It is clear that the linear stage model would gain strength if cognitive states could be incorporated in the model. As a first attempt, Sanders (1981, 1983) has proposed a model in which the processing stages are related to the three energetical supply systems of Pribram & McGuiness (1975). The arousal system provides the energetical supply for encoding, the activation system is thought to be connected to motor adjustment whereas effort would influence the choice stage. It should be noted that this model is a promising start but additional support is needed. Especially the evidence regarding the connection choice and effort is still quite meagre. The interconnections between the various energetical mechanisms make it hard to disentangle the loci of effect of the experimental manipulations. As yet the model can only incorporate four processing stages: the three stages mentioned above and stimulus preprocessing which may not require a separate energetical resource. Incorporating the other stages poses a dilemma: if each stage requires a seperate energetical supply this will lead again to inflation of the proposed energetical mechanisms and hence to inflation of the whole model.

Although the model is not without problem it can be an important step in information processing research. Different behavioral and physiological results and notions merge together into a model, and a cognitive concept of stress is put forward. Even more important is the attempt to examine the converging lines between the functional and structural approaches of human processing. Gopher and Sanders (1984) have shown that linear stage and resource volume models have more in common than originally thought. If the energetical supply to the stages can be considered as resources in the sense of resource volumes, attentional aspects can be incorporated in the linear stage model. If the amount of capacity is related to the amount of energetical supply to a particular stage, it is possible to find out which stages are selectively influenced by energetical state variables. Sanders (1983) has discussed the specific effects of suboptimal conditions on reaction time (e.g. Frowein, 1981b; Sanders, Wynen & v. Arkel, 1982: effects of amphetamine, barbiturates, and sleep state). In addition effects of cognitive states like Knowledge of Results, time pressure, and Time On Task can be analysed. Thus the AFM may not only be a tool for discovering computational processing stages (Sternberg, 1969), but might be also suited for the analysis of specific effects of resource allocation. The line of reasoning developed for stages is the same as that for resources: Whenever a variable affecting resource allocation is manipulated together with a variable which influences a computational stage, finding an interaction between these variables means that the stage which is influenced by the computational variable gets his energetical supply from the resource which is influenced by the state variable.

Although this model could be an important step towards integrating multiple resource and linear stage notions, Gopher & Sanders (1984) argue against efforts to develop a single experimental paradigm servicing both. The two approaches have different methodologies, different interests and can answer different questions. According to Gopher and Sanders (1984) converging evidence should be obtained along the lines of back-to-back experimentation.

CRITICISM OF THE AFM

The AFM is criticized in different ways (Hockey, 1979; Pieters, 1983; Prinz, 1972; Rabbitt, 1979; Stanovich & Pachella, 1977; Taylor, 1976). Taylor has claimed that it is logically possible that two variables both affect the same processing stage and yet show additive effects. An interaction could be masked if the two variables affect the stages in opposite ways (e.g. speed up and slow down). Pieters (1983) shows on logical grounds that a pattern of interactions is not sufficient for estimating the number of stages. Stanovich and Pachella (1977) argue that response selection will proceed in parallel with identification of the signal. Townsend (1976) has shown that, mathematically, parallel models are equivalent to serial models. These arguments are valid to some extent but it should be realized that models are not solely jugded on the basis of mathematical or logical arguments. Empirical evidence together with the most appropriate and parsimonious explanation, is at least equally relevant.

Other arguments against the AFM start off from a different rationale (Hockey, 1979; Rabbitt, 1979). The resource strategy approach claims that changes in task demands change the architecture of a processing sequence. It is obvious that this top-down approach rejects the use of the AFM. This resource strategy approach differs from the earlier discussed resource volume approach in that the amount of allocated resources change the nature of the reaction task. In contrast, resource volume models assume that the amount of output may change but not the kind of output (see also Section 5.4.2.). This is in line with the AFM which states that the duration of the stages can change but not the quality of the output.

Thus the resource strategy approach does not assume the task invariance assumption, whereas this assumption is an absolute prerequisite for applying the AFM. The nowadays popular research topic of controlled-automatic processing can be analysed with the AFM as long as the seriality of stages and task invariance are guaranteed. In most cases of automaticity, however, the structure of the task changes as a function of practice. This is no problem for the AFM as long as the changes are a matter of intra-stage change. The logic of the AFM prevents any interstage change. Therefore, topics such as task automaticity in relation to dual task performance should not be analysed in terms of the AFM. The next section will deal with the serial-parallel controversy with the aim of checking to what extent the stage framework may still apply.

SERIAL VS PARALLEL PROCESSING.

It is fair to say that a major theoretical issue in current cognitive psychology concerns the nature of human information processing as either serial or parallel. In the serial and parallel notions which are outlined here it is useful to define the term "element". This is the smallest unit of information processed in a particular stage of a particular model (Taylor, 1976). In a serial model each element is processed one at a time in a sequential order. Completion of one element initiates the processing of the next. In a strictly parallel model the processing of all elements is simultaneously initiated. The distinction between parallel and serial exhaustive models is illustrated in Figure 2.

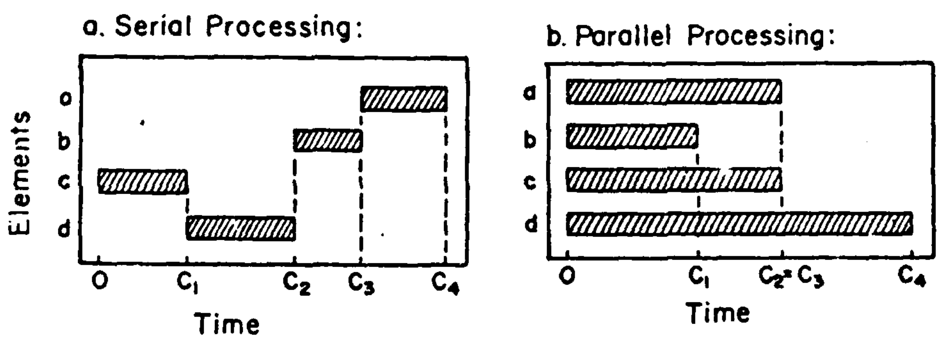


Figure 2: A representation of the serial or parallel exhaustive processing of four elements. (from Taylor, 1976)

Changing the processing duration of an element in the serial model changes the overall reaction time, whereas only an influence on element d, the critical element, will change the overall reaction time in the parallel model. It should be noted, however, that other elements can become critical. These two models are taken as representatives of the basic methods of processing. It would be more plausible to assume that information processing is a combination of serial and parallel processing. This last statement is not in disagreement with the linear stage approach, because parallel processing within stages is possible without violating the assumptions, since it does not change the interstage results. Futhermore, it is important to note that parallel processing between stages would ultimately lead to postulating a single stage. Hence, by virtue of the method itself, full parallel processing can never be considered with the AFM.

CHOICE REACTION AND PARALLEL PROCESSING

The number of alternatives and the relative SR frequency can give rise to parallel processing. Consider a standard lights and key situation with four alternative lights , and corresponding keys. The model of Hick (1952) assumes that subjects, in order to identify the locus of the stimulus, carry out two successive binary decisions. In a four choice condition the first binary decision step reveals that the stimulus is in the left pair, while the second step proceeds the preparation to respond to the left is already initiated. This situation suggests that there is a partial overlap of stimulus identification and response choice. This bias can be enlarged if there is one SR pair which is presented more often than other possible pairs. It should be realized however that this example is a rather hypothetical one and is only used to show possibilities for parallel processing.

Stanovich and Pachella (1977) have proposed a model in which the stages identification and response selection overlap. They argue that in a verbal naming task there is a two stage encoding process: first, feature extraction and second, identification with a subsequent feedback loop to feature extraction. They claim that a naming task requires an identification stage to get the actual name code, whereas this stage is deleted when a key press task is used. In a key press task the stimulus code (end product of feature analysis) directly determines the response code. Because of these parallel processes between stages, additive as well as non-additive effects can be found between signal contrast and SR compatibility. In the case of a highly overlearned response (naming a word) response selection is no longer involved while the identification stage becomes relatively dominant, and can be influenced by relative SR frequency. It seems that Stanovich and Pachella compare two different tasks. A very high SR compatibility leads to automatic response choice, in which case the task structure may indeed change and other models should be applied.

The distinction between name code and physical code suggests another area of informating processing where parallel processing has been assumed. Posner (1978) proposed a system of isolable processing, in which processing codes are operating in parallel and independent from each other. Presenting a letter to a subject will lead to the formation of two different codes, one representing the visual code of the letter and the other its phonetic recording. Both are representatives of the input. In principle one could argue for a serial process in which first a physical code is generated while a name code is made (as Stanovich & Pachella claim) when this code is not appropriate for the response. Posner (1978) shows with his matching experiments that both codes are used to achieve matches although physical codes are always faster. For example Thorson, Hockhaus and Stanner (1976) compared the effects of visual confusable items with acoustically confusable items in a successive matching experiment. When the interval between the two letters was short (shorter than 1.0sec) acoustic confusability had no effect on reaction time but at one second it began to produce a strong interference. On the other hand visual confusability showed a strong interference with short intervals. The basis of the match seems to change over the course of the interval. With short intervals the physical match dominates in speed and is already available. This produces an interference with visual but not with acoustic similar stimuli. The parallel "horse-race" model of Posner (1978) seems to account for the data (see also Sanders contribution).

It seems that these results clearly violate the AFM assumptions of seriality. But there is a way out: it can be assumed that within the encoding stage such parallel encoding processes take place. Perhaps the cleaning up of an degraded signal occurs also by means of a physical code. This way out is somewhat dangerous: Given the horse-race model of Posner, it is plausible to assume that under changing circumstances another horse will win, that is the name code may be available earlier then the physical code. Then the assumption of equal stage output is violated. It may be superfluous to state that the analysis of a name match with a physical match task can never be done by means of the AFM. Name versus physical code represents variations in tasks instead of variables.

Posner (1978) proposed a model in which "psychologic pathways" are central in encoding. He defines a pathway as "a set of internal codes and their interconnections that are activated automatically by presentation of the stimulus." This implies invariance between the input and the isolable systems. With regard to the name code-physical code discussion automaticity means that the codes are achieved without intention, without awareness and without interference with other ongoing activity. A well known example of this invariance is the Stroop color word test. Subjects want to avoid processing some aspects of the stimulus (the color) but it seems impossible to neglect the word. Furthermore Posner`s cost-benefit analysis (Posner & Snyder, 1975) shows the distinction between automatic parallel effects (no costs, only benefits) and effects of attentional mechanisms of limited capacity (costs and benefits). Given these considerations it is obvious that the suggestion of automatic and parallel processes within the encoding stage is not plausible. According to the energetical stage model (Sanders, 1983) this stage does not operate resource free and more importantly, according to Posner (1978), these automatic processes are supposed to occur in a very early stage of the reaction proces. Perhaps automatic processes are operating witin the stimulus preprocessing stage, which is considered to be resource free. Alternatively, a name and physical code can only be made after transport of sensory input (preprocessing stage) and after a general feature analysis (e.g. cleaning up the degradation) in the encoding stage. It can be assumed that the input for the parallel "code" processes have to be of a certain quality, which implies that, at least under some conditions, two perceptual stages precede "code" processing. This would mean that the automatic "code" processes occur within the identification stage. Hence, in conditions were degraded or unfamiliar signals are used the automatic parallel effects disappear (Sanders, 1983) because at least one resource-consuming encoding stage preceeds the identification stage.

It should be noted that Posner (1978) does not favor a linear stage model. He argues for a reaction time analysis without any assumptions regarding serial or parallel processing. His mental chronometry method reveals a great deal about the structure of internal processes, although it is clear that it is less powerful than the AFM.

The discussion serial or parallel processing can be further elaborated if the area of search tasks is considered (see also Section 5.4.5.). The well known classification task of Sternberg (1966) was reanalysed with the aid of the AFM (Sternberg, 1969). Sternberg (1969) concluded that at least four processing stages were involved: "stimulus encoding", "serial comparison", "binary decision", "translation and response organisation". The output of the encoding stage is send to the stage "serial comparison" whose duration depends linearly on the size of the memory set. Sternberg (1969) assumed a linear exhaustive process witin the comparison stage: for each member of the memory set a substage was postulated in which the representation of the teststimulus is compared with one of the members of the memory set. Relating these findings to the linear stage model of traditional choice reactions it is thought that an extra stage, "serial comparison", is included, which provides the information for response choice. Visual search experiments have shown that under specific conditions the linear relationship between the amount of comparisons and reaction time disappears (e.g. Neisser, 1963; Schneider & Shiffrin, 1977), suggesting that parallel processing occurs. Schneider and Shiffrin (1977) have argued that human performance is the result of two qualitatively different processes referred to as automatic and controlled processing. If target and non-target have remained fixed over trials (CM-stimuli) an automatic process can develop. Probably Sternberg (1966) did not use CM-stimuli, and a automatic parallel process could not evolve. Schiffrin and Schneider (1977) have shown with a Sternberg task that the linear relat onship between memory-set-size and reaction time disappeared under CM conditions although a small effect of memory-set-size remained.

Given the stage model of Sternberg it seems that the automatic mode of processing developed under CM conditions enables the serial search to be bypassed. Although this is an appropriate explanation, one has to consider reasons why this "bypassing" is possible. There are at least two possible explanations (Neumann, 1984). First, that which is automatic might be the parallel identification of all stimuli, at least up to the point where the attributes that specify the target become available. In Sternberg's model this might imply that only the target is send to the next stage and therefore comparisons are no longer neccessary. This would mean that parallel processing occurs within the perceptual stages of the linear stage model. Second, automatic might be an "automatic- atttention response" (Schneider & Shiffrin, 1977), which leads to attentional selection of only the target. This second explanation claims that non-targets are not processed, which again would mean that only a target is sent to the next stage. Although these explanations are logically independent a combination of both kinds of automaticity is also likely (Shiffrin & Schneider, 1977).

It can be argued that parallel perceptual processing in the sense of Schneider and Shiffrin can only develop if there is a strong S-R mapping, that is an attentional response is connected to particular target stimuli.

The invariance between a set of targets and a single response is thought to be crucial. It seems that it is not difficult to combine stimuli but only to deal with them independently at the same time. Neumann (1986) claims that capacity limits are related to limitations in processing stimuli without combining them. Serial processing might be considered as this limiting factor.

In case of an automatic process, an input automatically activates another process, that is processes occur as a passive consequence of stimulation. Automatic does not imply parallel processing but the reverse might be true: parallel processes often appear to be automatic. Situations in which the data-driven invariant connections between processes cannot develop, e.g. VM- stimuli, degraded perceptual quality or lack of practice would require a serial capacity consuming mode of processing.

In traditional choice tasks, processes within a stage need not be either "automatic" or "controlled". They can be automatic under some conditions- e.g. clearly visible letters and may be controlled in others- e.g. after degradation. Automatic and parallel processes might occur witin a stage but parallel processing between stages is rather hypothetical. Parallel processing within a stage cannot be analysed by means of the AFM, and in a traditional choice task it is rather speculative to assume parallel processing within stages. If, in a traditional choice task, parallel processing within stages would be plausible, one might predict more often violations of the assumption of constant stage output. If a task variable selectively influences one of the parallel processes within a stage the output of that stage will change.

SUMMARY.

Mental operations can function in two different modes (Posner & Snyder, 1975, Shiffrin & Schneider, 1977). Processes in the first mode occur as a passive consequence of stimulation and take place in a parallel capacity-ree manner. It is argued that an invariant connection between processes is a prerequisite for the occurence of parallel processing. The passive consequence is the result of this developed invariance. If these processes occur within the stages of the linear stage model and the strict assumptions of the AFM are not violated, the AFM can still be an appropriate tool, although it does not help to reveal the parallel aspects of the reaction process. Processes in the second mode are controlled by consious intentions, and are subject to capacity limitations. For this mode of processing the AFM is the most appropriate tool for analysis, but again violations of the assumptions are not allowed. Further research with regard to the linear stage model should focus on converging evidence regarding the now existing stages. It is important that the stages do not only represent a task variable but that they can be considered as psychological meaningful units. The cognitive states incorporated in the elaborated stage model could be of value for this purpose. Futhermore, the relations among the energetical supply mechanisms and the influence of different cognitive states deserve futher experimentation.

With regard to the parallel mode of processing it is concluded that parallel processing takes place as a passive consequence of stimulation, by nature these processes operate resource free. Finally, it is thought that parallel processing occurs on the perceptual side of the reaction process (see f.e. Kantowitz & Knight,1976; Posner & Boies, 1971; Schneider & Shiffrin, 1977). Serial processing is thought to be connected with processing limitations. It can work as a filter, that is the mere seriality garantees that interferences due to crosstalk between parallel processes cannot occur. Future research concerning serial or parallel processing should be applied to components of the reaction process instead of reaction process as a whole. It might be that automatic parallel processing within certain stages of the reaction process does not show up because the reaction as a whole disguises any parallel processing. Experimental paradigms which can disentangle the perceptual, decisional and motor components of the reaction process are in this respect of great importance. For example Sanders` functional visual field might be appropriate (Sanders & Houtmans, 1985). Given these possiblities one can study the nature of parallel processing, and find which factors do change the mode of processing.

REFERENCES.

Bertelson, P., & Barzeele J. (1965). Interaction of time uncertainty and relative signal frequency in determining choice reaction time. Journal of Experimental Psychology, 70, 448-451.

Broadbent, D. E., & Gregory, M. (1965). On the interaction of S-R compatibility with other variables affecting reaction time. British Journal of Psychology, 56, 61-67.

Donders, F. C. (1868). Die Schnelligkeit psychischer Prozesse. Archieve für Anatomie und Physiologie, 657-681.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. Journal of Experimental Psychology, 67, 103-112.

Fitts, P. M., Peterson, J. R., & Wolfe, G. (1963). Cognitive aspects of information processing. II. Adjustments to stimulus redundancy. Journal of Experimental Psychology, 65, 423-432.

Frowein, H. W. (1981a). Selective drug effects on information processing. Doctoral Thesis, Tilburg.

Frowein, H. W. (1981b). Selective effects of barbiburate and amphetamine on information processing and response execution. Acta Psychologica, 47, 105-115.

Gaillard, A. W. K. (1978). Slow brain potentials preceding task performance. Amsterdam: Academische Pers.

Gaillard, A. W. K. (1980). The use of taskvariables and brain potentials in assessment of cognitive inpairment. In B.M. Kulig et al. (Eds.), Epilepsy and behavior. Lisse, The Netherlands: Swets and Zeitlinger.

Gopher D., & Sanders A. F. (1984). S-Oh-R: Oh stages! Oh resources!. In W. Prinz and A. F. Sanders (Eds.), Cognition and Motor Processes. Heidelberg: Springer.

Hick, W. E. (1952). On the rate of gain of information. Quarterly Journal of Experimental Psychology, 4, 11-26.

Hockey, G. R. J. (1979). Stress and the components of skilled performance. In V. Hamilton and D. M. Warburton (Eds.), Human stress and cognition. New York: Wiley.

Holender D., & Bertelson P. (1975). Selective preparation and time uncertainty. Acta Psychologica, 45, 193-203.

Kantowitz B. H., & Knight J. L. (1976). Testing tapping time sharing II: Auditory secondary task. Acta Psychologica, 40, 343-362.

Miller J. D., & Pachella R. G. (1973). The locus of the stimulus probability effect. Journal of Experimental Psychology, 101, 227-231.

Neisser U. (1963). Decision time without reaction time: Experiments in visual scanning. American Journal of Psychology, 76, 376-385.

Neumann, O. (1984). Automatic processing: A rewiew of recent findings and a plea for an old theory. In W. Prinz and A. F. Sanders (Eds.), Cognition and Motor Processes. Berlin, Heidelberg: Springer-Verlag.

Neumann, O. (1986). ... In: H.Heuer and A.F. Sanders (Eds.), Issues in cognition and motor behavior. Hilldale, NJ. (in press).

Pachella, R. G. (1974). The interpretation of reaction time in information processing. In B. Kantowitz (Ed.), Tutorials in performance and cognition. Hillsdale, NJ: Erlbaum.

Pachella, R. G. & Fisher, D. F. (1969). Effects of stimulus degradation and stimulus similarity on the trade-off between speed and accuracy in absolute judgements. Journal of Experimental Psychology, 81, 7-9.

Pieters, J. P. M. (1983). Sternberg's additive factor method and underlying psychological processes: some theoretical considerations. Psychological Bulletin, 93, 411-426.

Posner, M. I. (1978). Chronometric explanations of mind. Hillsdale, NJ: Erlbaum.

Posner, M. I. & Boies, S. J. (1971). Components of attention. Psychological Rewiew, 78, 391-408.

Posner, M. I. & Snyder, C. R. R. (1975). Facilitation and inhibition in processing of signals. In P. M. A. Rabbitt and S. Dornic (Eds.), Attention and Perforance, 5. New York: Academic Press.

Posner, M. I., Klein R., Summers, J. & Buggie, S. (1973). On the selection of signals. Memory and Cognition, 1, 2-12.

Pribram, K. H., & McGuinness (1975). Arousal, activation and effort in the control of attention. Psychological Rewiew, 82, 116-149.

Prinz, W. (1972). Reaktionszeit - Fractionierung durch Varianzanalyse? Archiv für Psychologie, 124, 240-252.

Raab, D., Fehrer E., & Hershenson M. (1961). Visual reaction time and the Broca-Sulzer phenomenon, Journal of Experimental Psychology, 61, 193-199.

Rabbitt, P. M. A. (1979). Current paradigms and models in human information processing. In V. Hamilton and D.M. Warburton (Eds.), Human Stress and cognition. New York: Wiley.

Sanders, A. F. (1970). Some variables affecting the relation between relative signal frequency and CRT. In A.F. Sanders (Ed.), Attention and Performance, 3, 45-55 (Acta Psychologica 33).

Sanders, A. F. (1977). Structural and functional aspects of the reaction process. In S. Dornic (Ed.), Attention and Performance 6. Hillsdale, NJ: Erlbaum.

Sanders, A. F. (1980a). Stage analysis of the reaction processes. In G. E. Stelmach and J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland Publ.

Sanders, A. F. (1980b). Some effects of muscle tension on choice reaction time and movement time. In R. S. Nickerson (ed.), Attention and Performance 8, Hillsdale NJ: Erlbaum.

Sanders, A. F. (1981). Stress and human performance: a working model and some applications. In G. Salvendi and M. J. Smiths (eds.), Machine

pacing and occupational stress. London: Taylor and Francis.
Sanders, A. F. (1983). Towards a model of stress and human performance. Acta Psychologica, 53, 61-97.
Sanders, A. F., & Houtmans M. J. M. (1985). There is no central stimulus encoding during saccadic eye shifts. a case against general parallel processing notions. Acta Psychologica, 60, 323-338.
Sanders, A. F., Wijnen, J. L. C. and v. Arkel, A. E. (1982). An additive factor analysis of the effects of sleep-loss on reaction processes. Acta Psychologica, 51, 41-59.
Schneider, W., Dumais, S. T., & Shiffrin R. M. (1984). Automatic and control processing and attention. In R. Parasuraman and D.R. Davies (Eds.), Varieties of attention. New York: Academic Press.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Rewiew, 84, 1-66.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Rewiew, 84, 127-190.
Shwartz, S. P., Pomerantz, J.R. and Egeth, H. E. (1977). State and process limitations in information processing: An additive factor analysis. Journal of Experimental Psychology: Human Perception and Performance, 3, 402-410.
Spijkers, W. A. C. & Walter, A. (1985). Response processing stages in choice reactions. Acta Psychologica, 58, 191-204.
Stanovich, K. E. & Pachella, R. G. (1977). Encoding, stimulus response compatibility, and stages of processing. Journal of Experimental Psychology: Human Perception and Performance, 3, 411-421.
Sternberg, S. (1966). High-speed scanning in human memory. Science, 153, 652-654
Sternberg, S. (1969). The discovery of processing stages: Extentions of Donders` method. In W.G. Koster (Ed.), Attention and Performance 2, 276-315 (Acta Psychologica 30).
Taylor, D. A. (1976). Stage analysis of reaction time. Psychological Bulletin, 83, 161-191.
Theios, J. (1975). The components of response latency in simple human information processing tasks. In P. M. A. Rabbitt and S. Dornic (Eds.), Attention and Performance 5, 418-440. London: Academic Press.
Thorson, G., Hockhaus, L., & Stanners, R. F. (1976). Temporal changes in visual and acoustic codes in a letter matching task. Perception & Psychophysics, 19, 346-348.
Townsend, J.T. (1976). Serial and within-stage independent parallel model equivalance on minimum completion time. Journal of Mathematical Psychology, 14, 219-238.
Welford, A. T. (1980). Choice reaction time: Basic concepts. In A. T. Welford (Ed.), Reaction times. London: Academic Press.
Wertheim, A. H. (1979). Information processed in ocular pursuit. Thesis. Amsterdam: Academische Pers.

5.4.5. MEMORY SEARCH

by C. Hilka Wauschkuhn

The limitations of human memory are generally considered as major bottlenecks of performance. If at some point correct signals or responses are not available, performance stops altogether (Broadbent, 1984). Since this is particularly the case for various types of short-term memory demands, most performance assessment batteries include memory tasks, referring to the subject's ability to recognize previously presented items; some batteries include also tasks requiring short-term recall of most recent items (running memory) or of short lists (memory span). In all tasks items are kept in memory only for a relatively short period of time, expressed in seconds rather than in minutes.

The following pages will focus on short term memory (STM) and the process of item recognition, and in particular on variants of the Sternberg paradigm.

THE STERNBERG PARADIGM

Sternberg (1966) proposed an experimental paradigm, the so-called The task itself is simple and easy to perform; without time pressure error rates are minimal and even with time pressure they are usually low (1-2%). In contrast to most previous memory research paradigms, the measure of main interest is not failures of memory but time needed for successful recogniton. From a "stimulus ensemble" of all possible items a small number (1 to 6) of arbitrarily selected items is presented to the subject for memorization (positive set). Then a single test stimulus is presented and the subject as to decide whether or not the test stimulus is a member of the positive set by pressing an appropriate button. Reaction time (RT) is measured from stimulus onset to response. The interesting variable is the mean RT, usually plotted as a function of the positive set size. Apart from variations relating to factors like item quality (digits, letters, words, forms), size of the positive set, size of the test set, and probability of positive or negative responses, there are two major procedural versions of the paradigm: the varied set procedure where the subject memorizes a new positive set on each new trial and the fixed set procedure with one positive set for a whole series of trials.

Sternberg suggests that the outcome of the subject's memory search is the result of a fast serial scanning process, where the test stimulus is successively compared to each element of the positive set. This conclusion was based on the major findings in the basic experiments: mean RT is a linear function of the positive set size, the rate of search is about 40 ms/item when digits are used as stimuli, the slope is the same for positive and negative responses, and the zero intercept is about 400 ms.

While the slopes reflect memory search, the intercept reflects all other aspects of information processing. Sternberg (1975) has proposed a 4-processing-stage model where the intercept reflects (1) stimulus encoding, (2) a binary decision process (yes/no), and (3) translation and response processing.

On the basis of his data, Sternberg proposed two major characteristics of the search process. The scanning process seems to be exhaustive because of the parallel positive and negative latency functions. If the process were self terminating the mean rate of increase for positive responses should be about half the rate for negative responses, since all items must be checked before making a negative response whereas for positive responses an average of 50 percent scanning would be needed to arrive at a match (provided that serial positions are equally distributed).

At first sight, the assumption of exhaustive search may appear unreasonable and inefficient, but after a reanalysis of his experimental data as well as those from other investigators, Sternberg (1975) could corroberate this assumption. Thus he found a remarkable relation between the ratio of the slopes (positive vs. negative responses) and the scanning rate. Small ratios (i.e. positive responses much faster than negatives) were associated with a slow scanning speed, while 1:1 ratios were found with fast scanning. He argues that high scanning speeds are characteristic for exhaustive search. Once the search has started, it would be impossible or inefficient to interupt it somewhere in the middle when finding a match. Because there is time lost by checking for the occurence of a match after each comparison only with slower speeds self terminating search may develop. The scanning rate depends on the quality of stimulus material. Digits show the highest rate, followed by colors, letters, words, geometric shapes, random shapes and finally nonsense syllables (Cavanagh, 1972). Thus experiments with lists of words that are organized in categories (Naus, 1972) or with precueed recognition (Hendrikx, 1986) showed that search can at least be partially selective in the way that only the relevant subset is scanned.

Although scanning speed seems to be an essential feature in determining the nature of memory search the observed rage of speeds (38 to 90 ms/item) is much faster than covert speech (200 - 300 ms/item) (Landauer, 1962). Accordingly subjective reports indicate that memory scanning is not accessible to introspection (Sternberg,1966). This is very different from retrieval in serial recall tasks which actually take search rates of some 200 ms/item (Hendrikx, 1984).

The availability of items in memory was irrelevant in Sternberg's studies. Results obtained at varied and fixed set procedures are quite similiar although varied set items are supposed to be only stored in STM wheras fixed set items are probably additionally stored in LTM. Yet the data suggest that the same sort of memory is probably searched in both procedures. Sternberg suggests that prior to search LTM data are transferred into an active STM so that they are equally rapidly accessible. It should be noted though that more intensive practice with a fixed set shows pronounced effects on performance (Schneider & Shiffrin, 1977).

There are also a number of experimental results which the Sternberg model cannot easily explain. For example serial position effects have been found in various experiments which should not occur if search were really exhaustive. Again without further assumptions the model cannot explain the finding that repeated elements in the positive set as well as positive elements with a high probability show shorter RTs.

Sternberg (1975) has considered three alternative search models as candidates for explaining these effects.

SELF TERMINATING SEARCH

To reconcile probability effects on the one hand and parallel- linear set size functions on the other hand (in the fixed set paradigm), Theios (1973) proposed self-terminating search through the list containing all members of the stimulus ensemble coded in association with a response cue, positive as well as negative. The list is assumed to function as a push-down stack, which is searched until the probe is found. The members of the push-down stack are rearanged from trial to trial. The more recent or probable items tend to be on top of the stack.
Self-termination implies that minimum RT should be invariant with set size. Unfortunately minimum RT has been found to increase systematically with set size (Lively, 1972). Further an important shortcoming of this model is that, without additional assumptions about how negative items and their associated response cues are integrated in the list, it can not account for results from varied set procedures.

PARALLEL COMPARISONS

In parallel comparisons the probe is compared in parallel to all members of the positive set. It is assumed that all comparison processes share a fixed amount of processing capacity. This means that the more comparisons have to be made, the longer the decison will take.
Atkinson et al. (1969) and Townsend (1971) have assumed that comparison processes may start simultanuously with an exponentially distributed duration for single comparisons. When a comparison is completed his capacity is immediately redistributed to the other still active processes. Although there are some open points concerning the capacity concept, Sternberg (1975) concedes that this kind of model can explain the same scope of phenomena as his own serial comparison model.

DIRECT ACCESS

Trace strength models assume no search, but direct access to internal representations of the items. Members of the positive set acquire greater trace strength than nonmembers through rehearsal or presentation, serving as a discriminative signal for the later binary decision process. A functional relation between trace strength and RT is assumed.
There are different versions of the direct access hypothesis. For example, Corballis et al.(1972) and Nickerson (1972) proposed that for the most recently presented or rehearsed items trace strength is independent from set size. Like self-terminating search one implication of this assumption is that minimum RT for positive items is invariant with regard to set size. This is at odds with the results of Lively (1972). Baddeley & Ecob (1973) suggested that a fixed amount of trace strength is (unequally) divided between the elements of the positive set. This means there is less strength per item as the number of elements is larger, making discrimination of positive and negative items more difficult. Serial position effects can be accounted for by assuming that trace strength depends on the serial position of the item. To account for repetition effects it may be assumed

that the items gain available strength by multiple presentation in the same set. Assuming that trace strength is divided between all items in STM, additional load on STM should have a negative effect on RT. This prediction was not supported by the results of Darley, Klatzky, & Atkinson (1972).

There have been two suggestions to combine exhaustive scanning and direct access. One is to assume that serial priming followed by direct access is the basis of the recognition decision; the other is to consider recognition as the result of either search or direct access but never both.

Corballis (1975) suggested an integration between exhaustive scanning and direct access by considering the scanning process as a priming rather than a search process, - a simple activation of the stored representations of the items, followed by direct access. The exhaustive priming process can account for the form of the latency function, direct access and additionally priming effects for repetition and position effects. For repeated items multiple priming will result in shorter RTs; the combination of priming effects with effects of sensory activation or rehearsal may lead to positively inference as a basis of position effects.

Atkinson & Juola (1974) proposed a disjunct two-process model for LTM search which is fairly comparable to the Sternberg model for STM search.
When a list of items is learned their familiarity values are increased and additionally they are stored in an extra array. For extreme familiarity values later recognition is based on familiarity discrimination alone. This sort of response is fast but not perfectly accurate. For test stimuli with uncertain familiarity values an exhaustive search of the storage array is performed. Responses based on this kind of response mechanism start with a time delay of 70 ms but they are perfectly accurate and extremely fast (10 ms/item). They show similiar set size functions as the data from Sternberg tasks.
Applied to Sternberg tasks, the model will predict that decisions are usually made on the basis of the scanning process rather than on familiarity values, because as a rule there are no systematic differences in familiarty between positive and negative items. Serial position and repetition effects could be considered as exeptions. Because of their position or repetition some items may get higher familiarity values which allow the scanning process to be skipped and therefor lead to shorter RTs. Atkinson & Juola (1974) could support these assumptions.

VISUAL SEARCH AND MEMORY SEARCH

The basic features of the results on memory scanning (linear, and parallel set size functions with a slope of about 40 ms/item) are also be found in visual search (e.g. Atkinson,Holmgren, & Juola, 1972), where subjects memorize only one item and subsequently search for the presence or absence among several simultanuously presented items, under the provision that peripheral limits of search are well controlled (i.e. presentation of a limited number of items on a constant distance from a fixation point).

More recent studies have often combined aspects of visual and memory search. This research has shown more or less flat or curvilinear set size functions, in conditions with fixed sets and quite extensive practice, or in tasks in which positive and negative sets were categorically or physically distinct. In turn the traditional findings suggesting serial search were observed when the sets varied. In a combined visual search (4 items) and memory search (4 items) experiment self-terminating search regarding visual search prevails.

Schneider & Shiffrin (1977; and Shiffrin & Schneider, 1977) proposed that these different findings are due to two essentially different types of information processing; i.e. automatic and controlled processing. Based on their experimental findings, they proposed that automatic processing is generally fast, parallel, and not limited by short-term memory capacity. It seems fairly effortless and is not under direct subject control. It typically develops when subjects process stimuli in consistent fashion over many trials (fixed sets); once learned it is difficult to suppress, modify or ignore. Controlled processing is often slow, generally serial, effortful, and capacity limited. It can be controlled by the subject himself. It is needed in situations where the responses required to stimuli vary from trial to trial and is easily modified, suppressed, or ignored by the subject. Finally, all tasks are carried out by complex mixtures of controlled and automatic processes (Shiffrin & Schneider, 1977, 1984). Controlled processes are load dependent, automatic processes are independent or at least less dependent on load. Subjects control processes via allocation of attention, but that does not nescessarily mean they have insight into the nature of the ongoing processes. Set size dependency is associated with serial controlled search, set size independency with automatic detection.

Thus memory scanning in the sense of Sternberg is a controlled process. Shiffrin & Schneider accept Sternberg's interpretation that this scanning process operates as a serial and exhaustive process. Sternberg's failure to find different set size functions with fixed and varied sets is ascribed to the possibility that Sternberg has changed his fixed set too soon before subjects could develop automaticity. However, this does not solve the earlier mentioned problems.

Ryan (1983) has critically reviewed the automatic-controlled processing distinction and argued that this distinction has no theoretical value and is only a trivial redescription of the well known fact that in some cases performance is load dependent and in some other cases it is relatively load independent. Concerning item recognition he considers all defining features of the two processes to be at odds with various existing experimental data.

For instance he criticises (Schneider & Shiffrin (1985) concede that he is correct) that the theory can not explain that in a prememorized list paradigm, where subjects learn a list to criterion prior to testing, flattening of the set size function occcurs even on the first trials of experimental sessions, although flat set size functions are supposed to be indicators of automatic processing, which usually needs a considerable amount of practice to develop.
Schneider and Shiffrin (1985) responded that this paradigm cannot challenge their theory because on the one hand, learning and development of automaticity could have occured in the training phase, on the other hand

the paradigm allows many different mechanisms to operate (e.g. familiarity values (e.g. Mandler, 1980), or automatic detection and categorical classification (Shiffrin & Schneider, 1977), or controlled search in LTM), and offers no way to decide which processes were really used. They suggest that familiarity judgements are probably the basis of response in this paradigm.

Refering to an experiment of Forrin & Morin (1968) and a similiar one of his own, Ryan tries to demonstrate that two memorized sets can be scanned at once without costs, to question the essential assumption that two controlled processes can't occur in parallel (unless they don't stress STM capacity or run slowly and are sequentially interwoven).
Corballis (1986) showed that Ryan's argumentation is based on a logical error and in fact his data are more compatible with successive than concurrent scanning.

CONCLUSION

Many studies using variations of the Sternberg tasks have shown that the slope of response latency function is remarkably stable. Search rate changes systematically with different types of stimulus materials and appears to be inversely proportional to the STM capacity for the material in question (Cavanagh, 1972), and seems to be also independent of strategies or practice. Together this suggests that the Sternberg paradigm provides a more powerful measure for changes in memory performance than the usual capacity tests (Wickens, 1984; Smith & Langolf, 1981).

REFERENCES

Atkinson, R.C., Holmgren, J.E., & Juola, J.F. (1969). Processing time as influenced by the number of elements in a visual display. Perception and Psychophysics, 6, 321-327.

Atkinson, R.C., & Juola, J.F. (1974). Search and decision processes in recognitioon memory. In: D.H. Krantz, R.C. Atkinson, R.D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology. Vol I. (pp.243-93). San Francisco: W.H. Freeman.

Baddeley, A.D., & Ecob, J.R. 1973). Reaction time and short time memory: implications of repetition effects for the high-speed exhaustive scan hypothesis. Quarterly Journal of Experimental Psychology, 25, 229-240.

Broadbent, D.E. (1984). The Maltese cross: A new simplistic model of memory. The Behavioral and Brain Sciences, 7, 55-94.

Cavanagh, J.P. (1972). Relation between the immediate memory span and memory search rate. Psychological Review, 79, 525-530.

Corballis, M.C.(1975). Access to memory: An anlysis of recognition times. In: P.M.A. Rabbit & S.Dornic (Eds.), Attention and performance V (pp.591-612). New York: Academic Press.

Corballis, M.C. (1986). Memory scanning: Can subjects scan two sets at once? Psychological Review, 93, 113-114.

Corballis, M.C., Kirby, J., & Miller, A. (1972). Access to elements of a memorized list. Journal of Experimental Psychology, 98, 379-386.
Darley, C.F., Klatzky, R.L., & Atkinson, R.C. (1972). Effects of memory load on reaction time. Journal of Experimental Psychology, 96, 232-234.
Forrin, B., & Morin, R.E. (1969). Recognition times for items in short and long-term memory. Acta Psychologica, 30, 126-141.
Hall, J.F. (1983). Recall versus Recognition: A methodological note. Journal of Experimental Psychology, 9, 346-349.
Hendrikx, P. (1984). Temporal aspects of retrieval in short-term serial retention. Acta Psychologica, 57, 193-214.
Hendrikx, P. (1986). Compatibility of precueing and S-R mapping in choice reactions. Acta Psychologica, 62, 59-88.
Landauer, T.K. (1962). Rate of implicit speech. Perception and Motor Skills, 15, 646.
Lively, B.L. (1972). Speed/accuracy tradeoff and practice as determinants of stage durations in a memory search task. Journal of Experimental Psychology, 96, 97-103.
Mandler, G. (1980). Recognizing: The judgement of previous occurence. Psychological Review, 87, 252-271.
Naus, M.J., Glucksberg, S., & Ornstein, P.A. (1972). Taxonomic word categories and memory search. Cognitive Psychology, 3, 643-654.
Postman, L., & Rau, L. (1957). Retention as a function of the method of measurement. University of California Publications in Psychology, 8, 217-270.
Ryan, C. (1983). Reassessing the automaticity-control distinction: item recognition as a paradigm case. Psychological Review, 90, 171-178.
Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84, 1-66.
Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.
Shiffrin, R.M., & Schneider, W. (1984). Automatic and controlled processing revisited. Psychological Review, 91, 269-276.
Smith, P.J., & Langolf, G.D. (1981). The use of Sternberg's memory-scanning paradigm in assessing effects of chemical exposure. Human Factors, 23, 701-708.
Sternberg, S. (1969). High speed scanning in human memory. Science, 153, 652-654.
Sternberg, S. (1975). Memory scanning : New findings and current controversies. Quarterly Journal of Experimental Psychology, 27, 1-32
Theios, J., Smith, P.G., Haviland, S.E., Traupmann, J., & Moy, M.C. (1973). Memory scanning as a serial self-terminating process. Journal of Experimental Psychology, 97, 323-336.
Wickens, C.D. (1984). Engineering psychology and human performance. Columbus: Merill.

5.4.6. LEXICAL AND SEMANTIC ENCODING
by Andries F. Sanders

During the last decade a considerable amount of research has been devoted to the analysis of properties and aspects of lexical and semantic encoding. The general rationale is that perceptual processing--i.e. identification and integration of information—is embedded in a variety of memory systems. The combination of sensory input and memorial systems in the brain leads to meaningful interpretation of the environment and constitutes the basis for purposeful action. There is no perception without memory.

The most prevailing questions concerning lexical and semantic encoding center around the nature of codes —defined as the format by which information is represented (Posner, 1978)—, the complexity of codes - e.g. letters, words, pictures, sentences and still higher cognitive units-, the mutual relations between codes in the brain, which enables integration of the various aspects of percepts into meaningful units, and, finally, the ways of accessing codes.

A detailed treatment of theses topics covers quite a wide area of cognitive psychology, ranging from perceptual identification of simple signals to psycholinguistics and aspects of reasoning and problem solving. This is obviously not implied in the present outline, which will contain only a limited introductory sketch of the issues as mentioned above with regard to simple aspects of lexical and semantic encoding. There will be an emphasis on the results obtained with some major experimental paradigms, like letter matching, naming, lexical decision and priming.

THE NATURE OF CODES

With regard to establishing individual elementary codes--letters, simple shapes or pictures—much research has centered around Posner's lettermatching paradigm (e.g. Posner, 1970). One central outcome of this research is that the visual presentation of a letter or word independently activates a variety of codes on a physical "visual" level, a phonological articulatory level and on a more abstact category level. It has been consistently found that, when a decision can be made on the basis of a physical code—as in letter matching relating to physical identity—the reaction time is less than when a name code is required. There is fairly convincing evidence that a physical perceptual code and a phonological name code are established through parallel processing. A main empirical result in favour of this interpretation is that experimental variables affecting the "lower" physical code, do not or only marginally affect processing time when a "higher" level name code is required. If codes were serially established, a lower code should continue to have an effect on a higher level code. Another finding of interest is that categorical classification, say digits versus letters, can occur without first identifying the individual item on a name level. (e.g. Gleitman & Jonides, 1976; Jonides & Gleitman, 1976) Duncan has shown evidence that, irrespective of the final type of code - physical, phonological or categorical - the analysis of perceptual features of shape and form is the common base (Duncan, 1983).

It is interesting to note that the conclusions obtained from these predominantly "chronometric" experiments--response latency dependent on letter matching or naming—agree well with those that are based on traditional memory paradigms. Thus, the "levels of processing" view of Craik and Lockhart (1972) also suggests that an incoming stimulus is first processed in terms of its physical orthographic characteristics, then in terms of phonological characteristics, and finally, in terms of its meaning. In the original processing concept orthographic codes were assumed to be shallow and less durable; phonological codes were thought be somewhat "deeper", while a semantic code was supposed to be most durable. Although the originally serial concept of levels of processing does not appear to hold in " proper" memory studies either--consider for example the evidence that under proper circumstanes physical "visual" cues can lead to verbal recall of an item where verbal cues fail (Baddeley, 1978)--it is still current to conceive of the structure of memory as composed of a set of fairly independent domains (Baddeley, 1983). The domains are characterised by rich internal networks and connections and usually they correspond to different classes of cognitive or behavioral activity of which perception, reasoning and organisation of action appear to prevail (e.g. Morton, 1979).

It is quite evident that the classical distinction between layers of short and long term memory - e.g. Atkinson and Shiffrin, 1968) reappears in the domain theory, albeit more functionally connected to different classes of processing operations and with more emphasis on differentially structured systems of encoding. In contrast, differences in durability of short and long term traces are less stressed and certainly less conected to type of encoding. Thus, in the letter matching paradigm, differences in durability play at best a minor role. For instance, Kroll (1975) has suggested that, as such, visual letter codes are not less durable but that they are more susceptible to interference from subsequent visual items. Following the classical work of Shepard (1967) it has become clear that, that even when only briefly viewed, visual codes can be very resistent to forgetting in short-term recognition. This is especially valid when the visual codes consist of more richly structured pictorial displays.

One of the advantages of the letter-matching paradigm is that the temporal build-up of a visual code can be studied by way of the successive matching technique where the two letters are presented in succession. In that case the first letter can be encoded prior to the second letter, so as to save time in a subsequent same-different match. The first letter is ready for comparison at some interval, the size of which depends on the properties of the first letter—quality visibility etc.—as well as on the type of letter matching that is required. In other words on the "depth of processing". Various inferences about the properties of the codes, as derived from successive letter matching, are discussed in more detail by Posner (1978).

Yet, letter matching has the disadvantage that inferences about the build-up of the first letter—and hence about the established codes—can only occur on the basis of the final same-different response following presentation of the second letter. This means that there always remains confounding of encoding the first letter, the second letter, as well as the matching decision. Sanders and Houtmans (1985) have recently proposed a technique that enables separate measurement of the time needed to encode the first stimulus. In this technique the two signals are simultaneously presented under a wide horizontal visual angle. At presentation subjects fixate the signal that is presented at the left; this is followed by identification and a subsequent saccade to the signal presented at the right side. After encoding the right signal the trial is concluded by a same-different response. Sanders and Houtmans (1985) found that encoding the left signal is completed before shifting the eyes to the right signal. This renders the fixation time of the left signal a prime candidate for measuring temporal properties of encoding. The prospects of this technique in the analysis of perceptual and cognitive properties of semantic processing deserve further evaluation.

THE COMPLEXITY OF CODES

A major issue in lexical and semantic processing concerns the question about the relation between letter and wordperception and, in turn, how words are integrated into higher level propositional units under consideration of syntactic rules (e.g Kintsch, 1974). In this contribution only a selective outline is given of some recent trends on word perception following visual presentation.

First, then, there is evidence that the type of parallel build-up of physical (feature) and phonological (name) codes, as observed for single letters, also holds for letter strings. This seems to occur irrespective of whether the letters consist of unrelated strings or of related sequences that together constitute a word. In both cases physical identity matches are carried out faster than name identity matches. Meaningful words are only superior to unrelated letter strings in that the slope of the relation between reaction time and string size is considerably larger for unrelated strings (Eichelman, 1970). Again, when either words or non-words are presented and when, subsequently, one letter is tested by a forced-choice two alternative procedure, (correct, incorrect at a given serial position within the string) words do better than non-words (e.g. Reicher, 1969). It is interesting that similar effects have been obtained for pronouncable non-words (pseudowords) (Massaro and Klitzke, 1979). Furthermore the effects are most pronounced with clearly displayed high-contrast targets and they are relatively independent of contextual constraints (e.g. Johnston, 1978).

Most current notions on word perception have their origin in Selfridge (1959) "shouting demon" model. The various models share the assumption of various levels of processing, in which a level or processing stage is characterised by a set of mutually strongly related nodes. As such the feature level, the letter level and the word level are commonly distinguished. Highly activated nodes on one level activate corresponding nodes on a suppraordinate level, which ultimately leads to a single word or, at least, to a limited set of candidates. In a recent quantitative version (Mc Clelland & Rumelhart, 1981; Rumelhart & Mc Clelland, 1983) there are feedback loops from the higher to the lower levels so that processing is not limited to a forward "bottom-up" path but allows a "top-down" inquiry as well (see also Paap, Newsome, Mc Donald and Schvaneveldt, 1982). The feedback principle allows for post-hoc analysis of a word into its elementary letter constituents, so that the letter level can profit-- or be misguided! -- from the word level. This has obvious advantages when accounting for results that show better identification of individual letters in a degraded string when the string constitutes a word than when it consists of unrelated letters. The important finding that pseudowords have an advantage over "really" unrelated letter strings is explained by the further assumption that pseudowords activate nodes that correspond to real words, although the activation levels are lower since no conclusive end result is obtained. In a series of experiments Rumelhart and Mc Clelland (1983) have tested the predictions of a computer simulation of their model with regard to the effects of duration and timing of letters in a string on perceiving a simple letter in that string.

Despite the impressive evidence in favour of the Rumelhart and Mc Clelland model, there remain a number of basic queries which the model not seems to address. First, although a visual and a phoneme level are distinguished on the letter level, both of which affect the word level, the model connects the phoneme level to auditory input and the visual level to a visual input. The phoneme level can be reached by a visual input but only following visual letter analysis; this is at odds with the parallel build-up of physical and name codes as implied by the work of Posner (1978). In addition the model has little to say how visual and phonemic letter codes cooperate in establishing word perception. Finally the tests of the model mainly concern experiments on identification of individual letters within strings. This paradigm is quite different from the letter and word matching studies as reported by Posner and coworkers. It remains to be established to what extent the experimental paradigms show converging evidence or have their typical artefacts that limit generalisation. Whatever may be the case, both approaches as well as some others--e.g. Eriksen and Schulz (1979)--show strong evidence for parallel processing in handling visual information on the feature as well as on the letter level in word identification.

MUTUAL RELATIONS BETWEEN CODES

Lexical codes are not independently stored but show pronounced patterns of interrelaTions. This is the general conclusion from research on lexical decisions (Meyer & Schvaneveldt, 1971) and on sentence verification (Collins & Quillian, 1969). It extends to encoding categories, like letters and digits. (Sanders & Schroots, 1968). Cognitive categories can actually be defined by the property that relations within categories are stronger, easier to activate and more persistent than relations between categories. Thus, in a memory span task a string of items like TCS 582 is retained better than T5C852. In a lexical decision task, subjects decide whether a letter string presented as a target consists of a word or a non-word. A popular variant of the lexical decision task is to present first a prime stimulus which, after a brief interval, is followed by the actual target. Lexical decisions to words that are related to the prime--say, doctor - nurse--are typically faster than lexical decisions to words that follow a neutral unrelated prime stimulus. Similar effects are found when subjects are asked to name the stimulus.

Three major principles have been suggested to explain the data on primed lexical decisions. The first and original explanation is in terms of "spreading activation", which is thought to be a automatic consequence of word recognition and activation of its corresponding memory representation. Activation is supposed to spread automatically along the path ways of the memory network to nearby word representations (Collins & Loftus, 1975), so that a related target has already a preactivated memory representation which faciliates subsequent processing. It shold be noted that "relatedness" does not equal "associative value", since strong priming effects have been found for words that are no strong associates but follow logically in a sentence (e.g. Levi, 1981; Foss, 1982). It has been known since the classical work of Anne Treisman (1964) that linguistic factors are relevant in determining which next word is predicted by the context of a sentence. Such results argue against too simple an associative network interpretation of "spread of excitation" (see also Neumann, 1984). Another point of interest is that the automaticity of activation is probably not very strong, since priming effects are eliminated by a concurrent verbal task, which means that the effect is vulnerable to interference. (Hoffman & Mac Millan, 1985). Various authors have suggested that, perhaps complementary to automatic components, there could be additional attentional biasing of memory representations, that fit the context of the prime. There is some evidence that subjects can be instructed to expect some, rather than other, prime-target relations, the activation of which then produces facilitation even in the absence of semantic relatedness. Thus, variation of the proportion of related prime-target pairs affects the extent of facilitation. This could be considered as evidence--although not watertight--for attentional biasing of some prime-target pairs. As de Groot et al (1986) have noted, attentional effects should only be found at somewhat longer intervals between prime and target--exceeding, say, 250 msec--since the development of an attentional bias is presumably time-consuming. Again, if attentional biasing is considered in terms of reallocation of limited capacity resources to some rather than to other items in memory, then Posner's costs-benefit analysis (Posner and Snyder, 1975) should apply. Thus unexpected targets should have a relatively long processing time, as is the case with common effects of relative signal freguency imbalance on choice reaction time (e.g. Sanders, 1970).

The attentional theory did not fare particularly well in the recent work of de Groot et al (1986). Their study contains a systematic analysis of the effects of intervals between primes and targets. Although the facilitatory effects of priming are found to increase at longer intervals (more than 250 msec) the expected inhibitory effects on unrelated items failed to show up. It can be added that de Jong and Sanders (in press) also failed to find any effect of relative signal frequency on perceptual processing of the signals, as estimated by way of the visual field paradiqm, that was discussed earlier in this contribution. This result--which needs further analysis before it can be considered as well established--suggests that attentional biasing of perceptual codes, that was originally implied in Broadbent's (1971) response set, is either impossible or at least highly limited. Instead this result suggests that attentional presettting of certain items hardly affects the speed of signal identification but is exclusively related to response selection, and to programming and preparation of action.

ACCESSING THE CODES

A major issue in research on selective attention has related to the so-called early-late selection controversy concerning the locus of selectivity, early theory suggesting a precategorical locus and late theory suggesting a postcategorical locus. Acording to the former view lexical and semantic codes are only accessed after passing a selective filter system (Broadbent, 1958), while the latter type of theory has invariably defended that all stimuli impinging on a receptor surface are being processed in parallel to a categorical--although not necessarily a "conscious"--level. The attentional bottleneck concerns selection for action, not for perceptual identification (e.g. Duncan, 1980).

Since the original statement of both early and late selection theory, numerous experiments have been carried out which are partly farorable to either view. Apparently the situation is much more complex than originally envisaged, and it is likely there is more than one kind of selective attention (e.g. Keren, 1977). Thus it is fair to say that a generally valid early selection view is no longer tenable. As briefly discussed before, highly probable words in the context of a sentence or very frequently occurring words have a high probability of access, even if they are to be ignored. Yet, this can be handled by early selection theory, since it has always claimed to deal with information rather than with stimulation as a criterion for attentional selectivity (Broadbent, 1982).

There may be at least three major principles of interest with regard to early or late selection and, hence, to the question of accessing lexical and semantic codes. The first principle concerns <u>perceptual overload</u>. Evidence in favour of early selection of material on the basis of physical properties like ear of stimulation, color or shape of visual material, is particularly evident in cases where subjects are faced with continuing streams of information or at least with divided attention to various locations (e.g. Francolini & Egeth, 1980; Neisser, 1963; Kahneman & Henik, 1977; Noble & Sanders, 1981). There is considerable evidence that a proper physical selection criterion impoves the rate of search and detection of critical targets through excluding coding of irrelevant items.

A second principle concerns the extent of <u>categorisation and learned relations</u> with regard to targets and non-targets. The work of Shiffrin and Schneider (1977) on the development of automatic detection is a case in point. Despite the absence of a physical property for early selection, and despite the necessity of visual search, a consistent target set tends to overcome selective constraints. Neumann (1984) has correctly argued that such evidence does not necessarily imply complete encoding of all materials in the display in order to determine a criterion for selecting the adequate target. The usual notions of internal memory representations may serve the aim of identification as well as attention. Yet, there is fair evidence for parallel processing and direct access to lexical representations, at least as long as the total number of relevant items is small. Fisher (1982) has recently suggested that parallel access to internal codes may be limited to the capacity of a buffer in which a restricted set can be tested at the same time with regard to class membership (target/no target). Furthermore it is clear that when patterns are less practiced or when signals are degraded, lexical access tends to occur considerably slower, along the lines of Sternberg's (1975) earlier analysis. At the same time it should be clear that the rate of search in deciding about class membership is so fast--about 40 msec per item--that a deeper level of serial processing--e.g. through name codes--cannot be assumed. Yet, varying class membership, degraded perceptual quality and lack of practice all add to capacity demands, so that a general conclusion that "perceptual encoding is automatic" is an undue generalisation. (see also Jonides (1985) for a similar argument).

A further principle that has been stressed by Broadbent (1982) and by Kahneman and Treisman (1984) concerns the question whether subjects know - and hence can attempt focussing attention - where a relevant target will appear, or that they do not know and hence are dividing attention like in search. In particular in the visual domain trends towards early selection have been observed in conditions of divided attention. In the former case interference from nearby signals prevails--such as in the Stroop paradigm or in the Eriksen paradigm concerning the effect of adjacent letters (e.g Neumann, 1984).

There is also considerable evidence that when a target calls for action, the system tends to act much more as a single channel in that simultaneously appearing targets are not detected (Duncan, 1980) This can be equally well explained by early as by late selection theory albeit through different principles. Early selection theory might maintain that without demands for action, there is little processing of information and hence no need for selective attention. Late selection could maintain "forgetting" of other targets while acting on one. Even when actions are required there are notorious examples of exellent multiple task performance that all bear upon extended practice in the tasks involved (e.g Wickens, 1984). As a final remark it should be stressed that when different types of codes (e.g. visual vs. auditory) are activated, the probability of cross-talk between the systems is decreased, and lexical access of competing signals is facilitated.

CONCLUSIONS

In summary, this limited sketch attempts to draw some main lines on recent thinking about the multitude of codes on various processing levels as well as about their modes of interaction and interference in lexical encoding. One emerging conclusion is that there are multiple codes in pattern perception and memory which differ in nature and structure and which are differentially relevant under various circumstances. Trivial as this conclusion may seem, it still avoids a traditional levels-of-processing notion in which visual codes are most primitive and semantic codes most durable. Yet some codes are established faster than others. A second emerging notion is that various codes are activated in parallel; they may easily interfere in conditions of focussed attention, and turn out to be separable in conditions of divided attention. Finally, although components of lexical encoding appear to be automatically established, this does not imply that "some processing stages are automatic". Signal quality, overload and specific demands for action are relevant in deciding about the extent of automaticity. This follows Posner's (1978) argument that few mental processes are automatic as such. Controlled processing comes into action whenever the demands of the task require intervention.

REFERENCES

Atkinson, R.C.,& Shiffrin, R.M. (1968). Human memory a proposed system and its control processes. In: K.W. Spence and J.T. Spence (Eds): The psychology of learning and motivation, II. New York: Academic Press.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. Psychological Review, 85, 139-152.

Baddeley, A.D. (1982). Domains of recollection. Psychological Review, 89, 708-729.

Broadbent, D.E. (1958). Perception and Communication. London: Pergamon.

Broadbent, D.E. (1971). Decision and Stress. London: Academic Press.

Broadbent, D.E: (1982). Task combination and selective intake of information. Acta Psychologica, 50, 253-290.

Collins, A.M., & Quillian, M.R. (1970). Facilitating retrieval from semantic memory: The effect of repeating part of an inference. Acta Psychologica, 30, 304-314.

Collins, A.M., & Loftus, E.F. (1975). A spreading activation theory of semantic processing. Psychological Review, 82, 407-428.

Craik F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. Journal of verbal learning and verbal behaviour, 11, 671-684.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. Psychological Review, 87, 272-300.

Duncan, J. (1983). Perceptual selection based on alphanumeric class. Evidence from partial reports. Perception and Psychophysics, 33, 533-547.

Eickelmann, W.H. (1970). Familiarity effects in the simultaneous matching task. Journal of Experimental Psychology, 86, 275-282.

Eriksen C.I.W., & Schultz (1979). Information processing in visual search: a continuous flow conception and experimental results. Perception and Psychophysics, 25, 249-263.

Fisher, D.L. (1982). Limited channel model of automatic detection: Capacity and scanning in visual search. Psychological Review, 89, 708-729.

Foss, D.J.(1982). A discourse on semantic priming. Cognitive Psychology, 14, 590-607.

Francolini C.M., & Egeth, H.A. (1980). On the nonautomaticity of "automatic" activation: Evidence of selective seeing. Perception and Psychophysics, 27, 331-342.

Gleitman, H., & Jonides, J. (1976). The cost of categorisation in visual search: Incomplete processing of target and field items. Perception and Psychophysics, 20, 281-288.

de Groot, A.M.B., Thomassen, A.J.W.M., & Hudson P.T.W. (1986). Primed lexical decision: The effect of varying the stimulus-onset asynchrony of prime and target. Acta Psychologica,61, 17-36.

Hoffman, J.E., & MacMillan, F.W: (1985). Is semantic priming automatic In: M.I. Posner, & O.S.M. Marin, (Eds.), Attention and Performance XI. Hillsdale, NJ: Erlbaum.

Johnston, J.C. (1978). A test of the sophisticated guessing theory of word perception. Cognitive Psychology, 10, 122-153.

Jonides, J., Naveh-Benjamin, M., & Palmer, J. (1985). Assessing automaticity. Acta Psychologica, 60, 157-172.

Jonides, J., & Gleitman, H. (1976). The benefit of categorisation in visual search: Target location without identification. Perception and Psychophysics,20, 289-298.

de Jong, F., & Sanders, A.F. (1986). Relative signal frequency imbalance does not affect perceptual encoding in choice reactions. Acta Psychologica, 62, in press.

Kahneman, D., & Henik, A.M. (1977). Perceptual organisation and attention. In: M. Kubory and J.R. Pommerantz (Eds.), Perceptual Organisation. Hillsdale, NJ: Erlbaum.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In: R. Parasuraman, & D.R. Davies (Eds.), Varieties of Attention. New York: Academic Press.

Keren, G. (1976). Some considerations of two alledged kinds of selective attention. Journal of Experimental Psychology: General, 105, 349-374.

Kintsch, W. (1974). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.

Kroll, N.A. (1975). Visual short term memory. In: D. Deutsch, & J.A. Deutsch (Eds.), Short-term Memory. New York: Academic Press .

Levy, B.A. (1981). Interactive processing during reading. In: A.M. Lesgold, & C.A. Perfetti (Eds.), Interactive processes in reading. Hillsdale, NJ: Erlbaum.

Massaro, D.W., & Klitzke, D. (1979). The role of lateral masking and orthographic structure in letter and word recognition. Acta Psychologica, 43, 413-426.

Mc Clelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception Part I. An account of basic finding. Psychological Review, 88, 375-407.

Morton, J. (1979). Facilitation in word recognition. Experiments causing change in the logogen model. In: P.A. Kolers, W.E. Wrolstad, & H. Bouma (Eds.), Processing visible Language.New York: Plenum.

Neisser, U. (1963). Decision time without reaction time: Experiments in visual scanning. American Journal of Psychology, 76, 376-385.

Neumann, O. (1984). Automatic processing: a review of recent findings and a plea for an old theory. In: W. Prinz, & A.F. Sanders (Eds.), Cognition and Motorprocesses. Berlin: Springer.

Noble, M.E., & Sanders, A.F. (1981). Searching for traffic signals while engaged in compensatory traching. Human Factors, 22, 89-102.

Paap, K.R., Newsome, S.L. Mc Donald, J.E. and Schvaneveldt, R.W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. Psychological Review 89, 573-594.

Posner, M.I. (1970). On the relationship between letter names and superordinate categories. Quarterly Journal of Experimental Psychology, 22, 279-287.

Posner, M.I. (1978). Chronometric explorations of mind. Hillsdale: Erlbaum

Posner, M.I., & Snyder, C.A.R. (1975). Attention and congitive control. In: R.L. Solso (Ed.), Information processing and cognition: the Loyola Symposium. Hillsdale, NJ: Erlbaum.

Rumelhart, D.E., & Mc Clelland, J.L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. Psychological Review, 89, 60-94.

Reicher, G.M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 81, 274-280.

Sanders, A.F. (1970). Some variables affecting the relation between relative signal frequency and CRT. Acta Psychologica, 33, 45-55.

Sanders, A.F., & Houtmans, M.J.M. (1985). There is no central stimulus encoding during saccadic eye shifts: a case against general parallel processing notions. Acta Psychologica, 60, 323-338.

Sanders, A.F., & Schroots, J.J.F. (1968). Cognitive categories and memory span: The effect of temporal vs. categorical recall. Quarterly Journal of Experimental Psychology, 20, 373-379.

Schvaneveldt, R., & Meyer, D.E. (1973). Retrieval and comparison processes in semantic memory. In: S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press.

Selfridge, O.G. (1959). Pandemonium: a paradigm for learning. In: The mechanisation of thought processes. H.M. Stationary office London.

Shepard, R.M. (1967). Recognition memory for words, sentences and pictures. Journal of Verbal learning and Verbal Behavior, 6, 156-163.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing II: Perceptual learning, automatic attending and a general theory. Psychological Review, 84, 127-190.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. Quarterly Journal of Experimental Psychology, 27, 1-32.

Treisman, A.M. (1964). Verbal cues, language and meaning in selective attention. American Journal of Psychology, 77, 206-219.

Wickens, C.D. (1984). Processing resources in attention. In: R. Parasuraman, & D.R. Davies (Eds.), Varieties of Attention. Orlando, Florida: Academic Press.