Human Factors Report HFR-15

HUMAN FACTORS IN

COMMAND-AND-CONTROL

SYSTEM PROCUREMENT

MONASH UNIVERSITY HUMAN FACTORS GROUP

DECEMBER 1985

C.V. Copas, T.J. Triggs and J.G. Manton[*]

Department of Psychology

Monash University

Melbourne

AUSTRALIA

DTIC
S ELECTE D
JAN 2 8 1988
D

National Library of Australia card number and ISBN 0 86746 312 0

----
* Human Factors Group, Aeronautical Research Laboratories, Melbourne.

88 1 20035

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report is a review of various human factors engineering techniques that could be used at the early stages of a procurement program for computer based tactical information systems. The report also discusses the various stages of the program with regard to the prominent human factors issues that should be addressed. Some recommendations are made concerning the management of procedures for implementing human factors and some study topics are suggested for short term projects in a research institution. Two appendices are included. Appendix A deals with issues related to user involvement in the design process and Appendix B outlines some studies that have examined human factors issues in design programs. An annotated bibliography of papers used for this report has been prepared as a separate document.

Experiences during World War II highlighted the requirement for the equipment designer to take account of the human operator in the design and development of systems. Accordingly, human factors specifications were developed and have been refined over the years. These specifications are meant to bind the designer to produce a well engineered workplace for the human operator. These specifications describe the capabilities and limitations of the human and relate these factors to the design of controls, displays, workplace layouts and the physical enviornment.

The adequacy of the specifications has been severely criticised over recent years. The most significant objection within the present context is that they fail to resolve important human factors issues which arise during the design of computer based information systems. Existing human factors specifications are largely concerned with the hardware side of the man-machine interface and they fail to take account of software development and the flow of information between the human and the machine. There has been a growing awareness that a key factor in system performance is the ability of the operator to understand the information being presented so that he can make optimal use of the functions available to him. It is in this area of human cognitive performance that system design specifications are required along with guidance of a more conceptual nature.

The design and procurement process has a number of identifiable stages within it. All of these require human factors inputs. Many different labels are used for the various stages, with the earliest being related to concept and project design definition. If human factors considerations are not taken into account at these critical early stages then irrevocable and performance limiting features are likely to be built into the system. It is therefore necessary to adopt methods that will promote decisions relating to the human factors aspects that will enhance system performance at all phases of development, particularly the early stages.

This report attempts to provide guidance concerning the types of methods to use. We have reviewed two potentially useful techniques for developing human oriented specifications in modern systems design. The techniques of performance diagramming and modelling represent systematic ways of attempting to forecast the human component of system performance.

There were two main criteria by which performance diagrams were evaluated in this report:

(a) their applicability to early stages of design, and

(b) the extent to which they address cognitive aspects of behaviour.

Nine techniques were evaluated and it was found that the various techniques focus on different and sometimes overlapping aspects of cognitive behaviour. The techniques of job process charts (Section 2.9) and process control diagrams (Section 2.10) show the most potential for use in developing system specifications.

Eight performance models relating to human/machine system performance were reviewed for the report. An analysis of the theoretical acceptability of these models was made on the grounds of their validity, generality and parsimony. The criteria used to compare the models were pragamatic. They were:

(a) their applicability at early stages of design,

(b) their relevance towards human-computer interactions,

(c) their ability to stimulate design solutions,

(d) their capability to forecast personnel requirements of systems, and

(e) the extent to which they take account of team behaviour.

Once again it was concluded that the models tend to have different characteristics and that no general purpose model exists. NETMAN (Section 4.2.3), HOS (Section 4.3.1) and queuing models (Section 4.5) were identified as showing potential for being useful in system development.

A third method of forecasting the impact of a system upon human performance is through the use of subjective judgement (Chapter 5). Although this technique is widespread, it has received little attention in the literature. Three aspects of systems design were identified in which this method has figured prominently, viz:

(a) visual display design,

(b) maintainability estimates, and

(c) personnel forecasts.

It was concluded that the use of expert judgement is valuable in many design situations, although the technique needs to be applied as objectively as possible. Some suggestions for achieving such objectivity are made.

The forecasting methods of diagrams and models reviewed in this report constitute the technical aspect of the work. In addition, a discussion is presented of selected human factors issues (Chapter 6) which commonly arise during systems design for which the technique should offer some solutions. The issues include:

(a) person-machine allocation of function,

(b) the role of decision-aiding systems,

(c) the role of hypothesis and option generation,

(d) decision-making versus stereotyped behaviour,

(e) manual back-up for automated and semi-automated systems,

(f) effects of unreliable data,

(g) operator strategies and system design,

(h) individual differences and system design,

(i) degree of flexibility of systems,

(j) voice versus non-voice communication,

(k) graphical versus textual displays,

(l) team structure,

(m) training,

(n) *personnel forecasts.*

The discussion of the above issues was designed to orient the reader towards areas where diagrams and models have potential to solve design problems which are often neglected. Where possible, relevant findings in the literature have been condensed and some design guidelines provided.

Based upon the large amount of literature which was reviewed under the contract, this report concludes with selected recommendations (Chapter 7) which should assist human factors input during system procurement. The recommendations cover managerial (Section 7.2) and research topics (Section 7.3). The managerial topics prescribe steps that should be incorporated into the system development contract and thus enhance the discriminatory ability of the procurer to assess competing design proposals. In summary form, these recommendations are:

(i) Documentation should ensure that user requirements have been investigated at the planning stage of development.

(ii) Criteria of human performance need to be specified at the planning stages of design. The contractor should respond to the global system requirements specified in the development contract with a scenario which delineates the criteria which human performance must meet in order to maintain system effectiveness.

(iii) The contractor should formalise all methods of human factors evaluation and document the results in a clear fashion.

(iv) Expert opinion as a means of systems evaluation should be structured and well-documented.

(v) *Simple time-based analytical models and operational sequence diagrams should be derived from the mission scenario. The time constraints of functions and information requirements should at least be analysed.*

(vi) A document store which includes abstract models of the functioning of all systems in operation would facilitate system re-design and future system specification.

(vii) Computer-aided design techniques are a means of ensuring a systems approach to design.

(viii) Systems should not be evaluated through the use of prototypes and mock-ups alone. *These techniques do have the advantage of permitting detailed evaluation of alternative system configurations and should be utilised in a comprehensive manner.*

(ix) The characteristics of the available personnel resource, namely numbers and level of training, should be a contractual specification which constrains the design. The contractor should provide documentation to ensure that those limits are not exceeded.

(x) If a building-block approach to design is followed, sub-system integration should be demonstrated.

The recommendations concerning the topics for research are appropriate for a short term research program conducted by a Research and Development agency. The topics identified are:

(a) Identify the facilities required by an operator for him to assume control of the system when automated functions fail.

(b) Produce guidelines on how to specify the degree of flexibility required in a system.

(c) Develop a method of predicting operator strategies in future system operation and how to design systems to take advantage of them.

(d) Identify the benefits of graphic displays over totes and give guidelines as to their design.

(e) Produce guidelines concerning the allocation of tasks between *members of a team.*

GLOSSARY

AI	:	artificial intelligence
APS	:	Analytical Profile System
ARI	:	(U.S.) Army Research Institute
ASW	:	anti-submarine warfare
Bottom-up	:	refers both to a form of modelling (in which total performance is found by aggregating less molar performance), and to a form of systems design (in which a number of sub-systems are combined).
c^2	:	command-and-control; the process thereof
c^3	:	command-control-and-communication; usually referring to an actual system
c^3I	:	command-control-communication-and-intelligence system
CAFES	:	Computer Aided Function Allocation and Evaluation System
COMCON	:	Command and Control System Simulation
CRT	:	cathode ray tube
DEI	:	Display Evaluaton Index
Development cycle	:	the formal stages through which military system design progresses
DQM	:	Decision Quality Matric
DT&E	:	developmental testing and evaluation
Heuristics	:	strategies or 'rules of thumb'
HFE	:	human factors engineering
HOS	:	the Human Operator Simulator model
IDEF	:	Integrated Computer Aided Manufacturing Definition
LAMPS	:	Light Airborne Multi-Purpose System
LSO	:	Landing Safety Officer
MAU	:	multi-attribute utility
Mil-Specs	:	military specifications (which form a contractual obligation during system design)
MMI	:	man-machine interface

NASA	:	National Aeronautics and Space Administration (U.S.)
OCM	:	optimal control model
OT&E	:	operational testing and evaluation
PROCRU	:	*Procedure Oriented Crew Model*
RAN	:	Royal Australian Navy
RN	:	Royal Navy
ROSIE	:	Rule Oriented System for Implementing Expertise
RPV	:	*remotely piloted vehicle*
SADT:	:	Structured Analysis and Design Technique
SAINT	:	Systems Analysis of an Integration Network of Tasks
SHOR	:	stimulus - hypothesis - options - response
STAMMER	:	*System for Tactical Assessment of Multi-Source Messages, Even Radar*
TACO	:	Tactical Air Co-Ordinator
Taxonomy	:	classificatory scheme
Technical determinism	:	refers to a form of design in which non-hardware variables (such as human factors) are neglected
Top-down	:	refers both to a form of modelling (in which performance is described as that which is sufficient to achieve system goals), and to a form of systems design (in which all design is a response to system goals and is organised hierarchically)
TOS	:	Tactical Operations Systems
USAF	:	U.S. Air Force
USCG	:	U.S. Coast Guard
VDU	:	visual display unit

LIST OF FIGURES

FOREWORD

The military commander requires the system under his command to perform a series of functions that will allow him to control the battlefield and defeat his adversary. The battlefield is competitive and the possession of sophisticated hardware allows the commander to approach his military goals from directions that his opponent may not expect. Western military weaponry generally relies on its sophistication to close accurately on its target. This requirement of precision combined with the additional cost of more sophisticated weapons means that the management of limited resources is an important issue in person-machine system design and operation. Furthermore, the speed with which modern warfare events occur and the greater distances over which control can be exerted, means that there is an additional need to produce a capable command information system.

The problems associated with implementing such a system stem from the commander's need to manage the resources available, each with its own inherent technological restrictions, given the objectives he wishes to attain. Each person responsible for a system component is in principle responsible for satisfying the goals of his immediate superior. If these goals could be defined precisely enough, then automated systems alone might achieve them. However, in the competitive environment of war the goals may change over time and may require equipment to be used in an unforeseen manner.

It is against this background that the requirements of the human operator must be taken into account at the earliest stages of the system procurement cycle. If human factors advice is not sought until after the concept, feasibility and product definition stages are over then its impact is likely to be minor or transitory.

This report provides a review of human factors techniques that could be used in the earliest stages of system design. One consideration was that it might be possible to supply the Service Staff Officer on a procurement project with a presentation of the techniques from which he could implement the human factor inputs for his current project. However, it became apparent, from the design techniques reviewed, that specialist background knowledge and experience is generally required (see Appendix A and B). The Staff Officer may therefore be advised to read the Executive Summary, Chapter 6 on Human Factors Issues in Systems Development and Chapter 7 that contains the recommendations for the management of human factors activities and some research topics. Only those who need more detailed information about system description should read Chapter 2, while Chapters 3, 4 and 5 provide specialized details about system models and evaluation.

Finally, it will be apparent from the unclassified bibliography that most of the techniques dealt with were developed and used by the United States Military and its contractors. This is almost solely due to a Congressional requirement that military projects must involve a human factors component. Relatively few references are made to European and civilian developments.

1. INTRODUCTION

1.1 Background

The major topic of this report is a review of various human factors engineering techniques that can be used at the early stages of a procurement program for computer-based tactical information systems. The report also raises some human factors issues that should be addressed during that procurement program. Some recommendations are made concerning the management of procedures for implementing human factors, and some topics for further research are identified.

Experience during World War II highlighted the requirement for the equipment designer to take account of the human operator in the design and development of systems. Meister and Rabideau (1965) indicated that the concern centred on the ability of the operator to use the new complex equipment both effectively, as a fighting machine, and safely. Whereas traditionally the artisan had always developed his equipment on a functional and evolutionary basis the designer of the newer mechanical hardware was somewhat removed from the environment where it was to be used. A systematic approach was then required. The seeds of awareness that human factors design inputs could have substantial effects on overall system performance was the start of the development of a systems approach to design (Singleton, 1974). As we have increased our ability to develop large scale systems so the design process has become more complex. It has become necessary to attempt to predict at the conceptual stage the way systems will actually perform. The relevant factors that contribute to system effectiveness need to be integrated into the design process.

Just as neglect of one of the various branches of engineering may compromise the design, failure to take account of human factors may decrease system performance. Carlson (1983) pointed out the operational limitations on the recovery of a LAMPS Mark III helicopter onboard US Navy frigates because of the poor design of the Landing Safety Officers control station. The problems were partially overcome by a costly redesign exercise. This process followed established analysis techniques that should have been used in the initial design process.

Following World War II a great deal of effort was directed towards developing specifications for the design of military equipment. Such specifications describe the capabilities and limitations of the human operator and relate these factors to the design of controls, displays, workspace layouts and the physical environment. Many of the principles are extensively documented in textbooks on the topic such as McCormick (1976); Morgan, Cook, Chapanis and Lund (1963); Van Cott and Kinkade (1972) and Woodson (1981).

Within the design of computer-based Command-and-Control (C^2) systems, there is a growing awareness that a key factor in system performance is the ability of the operator to understand the information being presented so that optimal use may be made of the functions that are available. This consideration does not arise in the design of conventional hardware-only systems. As a result, a technically determined approach to design is now seen as unsatisfactory (Howie, 1978), i.e. in which technical matters receive emphasis to the exclusion of other issues. Whilst hardware cost and availability may be major concerns during the development of large military systems, the advent of

computerisation has been responsible for increased concern for both human factors and software development.

Predictably, the adequacy of traditional military specifications in the purchase of complex systems has been criticised in recent times, e.g. U.S. General Accounting Office (1981). The most significant objection within the present context is that there has been a failure to resolve important human factors issues which arise during the design of computer-based information systems. Existing human factors specifications are largely concerned with hardware design, due to their emphasis on the physiological and anthropometric characteristics of the human. The specifications tend to neglect the cognitive or information processing characteristics of the operator, and thus provide little guidance (or constraint) upon software development. This is not to say that software design has been neglected entirely, but general specifications such as "interactive systems should be easy to use" have negligible practical significance. In short, whilst existing specifications may facilitate the design of keyboards and visual display units (VDU's), they do not provide guidelines for the design of features which distinguish a C^2 system (information based) from a weapon system (technology based).

The recent scrutiny of human factors specification methods has culminated in general agreement within the U.S. military that revised design guidelines are necessary. In fact, each of the three U.S. services recently commissioned research aimed at the preparation of a handbook for computer-based system design. The Navy work was carried out by Ramsey and Atwood (1979), the Army by Parrish, Gate and Munger (1981 a and b) and by Sawyer, Fiorello, Kidd and Price (1981), and the Air Force by Smith (1980, 1981).

Possibly the strongest common theme which has emerged from the work of these various investigations is the need for early human factors input to the design process of computer-based systems. The design and procurement process contains a number of identifiable stages (that are discussed more fully in Section 1.2). Many different labels have been applied to the various stages. However, the initial ones in system development are related to concept and project design definition. If human factors are not considered at this critical stage, then performance limiting features may be actually built into the system. In such an instance, an unnecessary burden is then placed upon training resources to maintain system effectiveness after the system becomes operational. Alternatively, experience with the operational system may lead to system re-design, which is usually a costly and inconvenient process.

The need for early human factors input to computer system development is well-expressed by Ramsey and Atwood (1979), viz:

> "In the design of interactive computer systems, virtually every decision which affects the functional behaviour of the system has direct human factors overtones. This claim can be made in the cases of automobiles, aircraft, and radios, too, but only in a much weaker sense. In a system whose basic function is communication with a user, and whose basic purpose often is to assist the user with tasks which are cognitive or informational in nature, human factors issues pervade the entire design process." (P.6).

The need for early human factors input has been a major factor in the move away from the systems specification of the traditional engineering type. As the issues which arise at preliminary stages of design tend to be somewhat intangible, design guidance of a more conceptual nature is now seen as necessary. Traditional human factors specifications are meant to be consulted after a particular design problem has been reached, whereas the more recent philosophy appears to be that guidance at all stages of design is necessary and that handbooks should actually anticipate design issues. The task of the guideline writer is therefore to distil and edit a widespread expertise that is largely undocumented.

While it would be desirable to provide data about the impact of system design upon human performance (particuarly if cognitive performance is addressed), there are more immediate concerns in the preliminary stages of design. Both designer and procurer alike need to be provided with tools that facilitate their definition of the system concept and permit an analysis of that concept. Such tools should allow an evaluation to be made of the preliminary design and of alternative designs.

This report has focussed on methods of human performance forecasting as the primary technical topic. These methods may be broadly classified as performance diagramming, modelling and methods of utilising subjective judgement. Depending upon the manner in which the methods are applied, they may be used to assist concept definition, system analysis and/or evaluation. The remaining content of this report is seen as complementing the review of these human factors methods, and is described more fully in Section 1.3.

1.2 Stages of System Design and Procurement

One of the features of the systems approach is the rigour with which it has become necessary to manage the design process (Singleton, 1974). In fact, it is almost a truism to say that the design process of complex systems passes through recognisable, sequential stages. Within the U.S. military procurement cycle, there are standard requirements for documentation at each of those stages. This documentation provides some degree of assurance that the contractor has addressed certain well-defined issues that arise during the development cycle.

The terminology that is used to describe the stages of design varies somewhat from author to author, but usually there is reasonable correspondence. Meister (1982a) has classified these stages as system planning, predesign, detail design, production, test and evaluation. Woodson (1981) prefers to name the sequence as concept formation, preliminary design, detailed design, prototype development and test, design modification, and production/delivery. The common elements uniting the two descriptions are these : first, systems must be planned (i.e., the goals of the system must be specified in order that the means of achieving those goals may be designed); next, the actual design proceeds from a conceptual to a more detailed form; lastly, the system must be evaluated to ensure its effectiveness before it is delivered to the customer.

Whilst these phases form distinct categories, they overlap in practice. That is, design tends to be an iterative process. For example, the original plan may be subject to modification as details of the design allow a clearer perception of the feasibility of the system concept.

Furthermore, system evaluation is rarely carried out just before the operational stage alone. There is often a contractual requirement for evaluation at all stages of design; an evaluation which becomes more precise as development proceeds. Evaluation at intermediate stages of design naturally leads to re-design in order that the system may better attain its objectives.

It is also a characteristic of computer based systems that they may often be modified after the formal design stages have ceased. This facility is largely due to the relative ease with which software may be re-configured, in contrast to hardware. This flexibility permits a system to be developed in stages, i.e. part of the system may become operational before the next part is introduced. An example of evolutionary design is where a manual system has one of its subsystems computerized as a forerunner to the total system being computerised.

Evolutionary design is said to have a number of beneficial human factors aspects (Eason, 1982). Firstly, the fact that experience with the system is gained in an incremental fashion promotes more thorough critical reviews. Secondly, the time constraints within which human factors inputs to the design process must occur are expanded, thus permitting more complete analysis. Lastly, user involvement is enhanced (see Appendix A), as the operators are in a position to criticise the system authoritatively when newer designs are considered.

Evolutionary design allows the goals of the system to remain flexible to some extent, which is desirable (Carley, 1967). This point is particularly relevant for C^2 design because those systems are frequently used for the performance of ill-defined tasks and may not, in principle, be amenable to an a priori method of design (Nickerson et al, 1977).

All of these exceptions from the 'ideal' development cycle can and do apply to C^2 procurement. However, this does not contradict the point that recognizable phases of design exist. Correspondingly, the type of human factors input that is required changes at each phase of design. Adopting the terminology of Meister (1982a), a brief discussion will follow of what that input should be in relation to some of the stages of the development cycle.

1.2.1 System planning

System planning initially requires that the purpose and objectives of the proposed system are defined. It is a feature of the systems approach that mission goals (in a military context) should be clearly articulated in order that the means for achieving those goals may be devised. An important part of this process of definition is that the criteria of system performance should emerge, against which later evaluative tests will be made (Woodson, 1981).

Within computer-based systems, the definition of the requirements of the operator is a major task at the planning phase of design. That is, the goals of the system should not be defined in an a priori fashion, but should be decided in conjunction with the needs of the proposed users of the system.

In commercial systems, a problem can arise where users agitate for goals that are irrelevant to, or that conflict with, those of the

total enterprise. This is not to say that user dissatisfaction is not to be regarded as a significant cost. In military systems, on the other hand, it is more likely that comprehensiveness of goal definition will serve a very pragmatic purpose, namely, the enhancement of the human contribution to system effectiveness. For example, if a C^2 system is about to be converted from a manual to a computer-based model, it is logical that the needs of the command first be investigated, in order to decide what the system should actually achieve. Neglect of the requirements of the command may lead to the introduction of an unwanted or inefficient system.

1.2.2 Predesign

The distinction between system planning and predesign is not clear-cut. System planning in this report has been described as a phase of goal definition alone; however, this phase tends to be closely linked with that of determining the means of achieving those goals. During this latter phase, system functions must be specified.

System functions are those activities which mission analysis has revealed as necessary for achieving system goals. For example, a weapons system as a minimum involves the functions of detection, tracking and weapons assignment. It is possible to conceive of systems containing a hierarchy of functions, ranging from the most global to the molecular. Those functions just quoted as an example constitute a global level of description. Each individual function may be further analyzed until a point is reached at which the tasks of a single operator are being described.

In hardware systems design, the major concern at the predesign phase is the allocation of function between people and machines. That is, as it is presumed that system functions are independent of each other to some extent, a reasonable issue is which of those functions should best be automated. It should be noted that such a design philosophy has not escaped criticism, and further details may be found in Section 6.2.

An important feature of the systems approach to design is that the system should first be conceived in purely functional terms (Singleton, 1974), i.e., the description of functions should precede the description of the concrete realization of those functions. Premature allocation of function tends to reduce the consideration of alternative system configurations, which inhibits the design process. In particular, technical determinism may constrain design methodologies such that automation is assigned to all functions for which it is possible to do so at a reasonable cost. Another fallacy may be the uncritical assignment of 'traditional' human functions. Further details may be found in Section 6.2.

Within computer-based systems, the allocation of function issue is even less straightforward. Because it is presumed that human and computer will co-operate on most tasks within interactive systems (by definition), it is no longer meaningful to speak about the allocation of function to human or machine. Rather, the essential design problem is to define the role of the human in the system. The most immediate decisions are those pertaining to operator participation and autonomy. Such decisions have a very pervasive influence on system development.

For example, decision support systems, such as those of C^2, may be designed in a variety of ways according to the implicit model of the role of the human. At one extreme, the computer may initiate all transactions to which the user must respond by selecting function keys; at another extreme, the user may manipulate the system directly through self-generated commands. Design issues such as whether to implement a decision-aiding system, and what form that aid should take, are also founded on some notion of the human's role. (See Section 6.3 for further details).

Software engineering practices may also involve a form of "technical determinism". Without a proper concern for user requirements, the design may cause the human to be 'locked out' of the system to some degree and hence the operator may be relatively limited in his ability to intervene. For example, the identification of friend from foe is an urgent matter during congested warfare conditions. It is possible to envisage a radar system automatically assigning status codes to foreign objects, while at the same time the command possesses further background information. In such circumstances, a facility for overriding, or modifying, the automated system is obviously desirable. Generally, the human needs to be given a central role in computer-based systems. Wohl (1982) has drawn similar conclusions in his review of human factors input to the Apollo space program.

One means of compensating for the negative effects of automation upon human performance, as already implied, is through the use of flexible systems. In more technical terms, systems in which the function of the human changes dynamically may have a number of advantages, e.g. Rouse (1977), particularly if the tasks which the system must accomplish are variable, or if different types of operators use the system. Further details may be found in Section 6.9. A prominent issue at predesign phases of development then becomes the basis upon which dynamic allocation of function should occur.

1.2.3 Detail design

Possibly the role that is most characteristic of human engineering in this phase is assisting the design of 'human-machine interfaces'. In general terms, interfaces are those parts of the system at which the human and the machine have the closest physical or informational contact. Interfaces may be regarded as the parts of the system that deliver information to the operator and through which the operator exerts actions. Displays, dials, controls and key boards are possibly the most common hardware items involved. Human engineering standards for the design of these items are relatively well-developed, and may assist the design or evaluation of these aspects of the system. Other aspects of the system for which human factors input is required are workspace layouts and the physical environment.

Within the software domain, design also proceeds from a conceptual to a more detailed form. As much of the human-computer interaction may be regarded as a dialogue, decisions about the content of that dialogue tend to precede the more detailed considerations about the form of that dialogue (Stewart, 1976). In fact, adopting the terms of linguistic analysis, Sibert (1983) has distinguished the semantic syntactical and levels of software design. Generally, the more abstract software decisions are concerned with the role of the human in the system,

while the more concrete decisions occur at advanced stages of design and are concerned with matters of syntax and format.

There is also likely to be a requirement for forecasts of the personnel characteristics of the human component of the proposed system. That is, both personnel numbers and training level need to be specified. These forecasts have the purpose of yielding preparatory information. Additionally, the forecasts may provide a check on the development contract if personnel resource specifications have been included.

Two infrequently discussed human factors issues at the latter phases of system development are the design of procedures and organisational structure. Whilst system design constrains the procedures that will be required for operations, it is possible that some tasks can be performed in a variety of ways. There is therefore a need for procedural analysis, particularly if an operational manual for the system is required. Procedural design may also compensate in part for poor system design, in much the same way as training programs may (see Section 6.9). Similarly, the organisational structure of human operators may be re-arranged, for example in order to equalise the workload (see Section 6.13).

Question of procedural design and team structures are often placed in the category of 'on the job' organisational problems, but in principle they are issues which could be anticipated at the earlier stages of design.

1.3 Scope of the Report

As discussed previously, the state-of-the-art in system design is such that the most critical issues arise at the early phases of development. Abstract issues such as concept definition require detailed attention at this stage. As a result, possibly the two most neglected areas of human factors input are requirements definition and the identification of the role of the human within the system.

The use of applied human factors methods is widely advocated as a means of assisting design at preliminary stages. There is a need for engineers to define the system concept, then analyze and evaluate the impact of their designs upon human performance. Accordingly, the use of performance modelling and diagramming (and, to a lesser extent, the use of subjective judgement) will be highlighted as design tools. The techniques of modelling and diagramming are well-established as a means of systems analysis within conventional engineering, as should their counterparts be within human factors engineering. The techniques may assist systems evaluation by providing quantitative forecasts of performance, that may be compared with criterion values. Requirements definition also necessitates some understanding of the tasks which the potential users of the system will perform, and modelling may be one means of acquiring that understanding.

The bias in the subject matter of this report should not be taken to imply that mock-ups, prototypes and hardware simulators have a minor role within systems evaluation. On the contrary, such techniques are probably essential for the evaluation of complex systems, and in many cases, would form a contractual obligation (as would evaluation of the operational system). However, in keeping with the policy that interactive

systems development requires early human factors input, methods of evaluation which may be applied before the detailed design stages have been given priority. (A critique of the relative merits of diagramming/modelling vs. mock-up evaluation may be found in Section 3.1.1.)

Appendix A also touches on the rather complex topic of requirements definition. There is a reasonably broad discussion of user involvement in systems design, including methods that may be used to promote user involvement. This topic at the least requires a further discussion of survey methodology in order to be comprehensive, for example, Ramsey and Atwood (1979), and Nadler (1981), and hence, it has been added as an appendix to the main work.

A second emphasized topic in this report is that involving a discussion of a number of human factors issues that may arise during system development. These issues are listed in Table 1. At one level, this discussion is designed to bring to the attention of designers some issues which might otherwise be neglected. At another level, an attempt has been made to relate these issues to the methods of systems evaluation that form the nucleus of this report. Human factors issues, if recognized, may implicitly lead to a design 'problem'. As a result, a decision must be made regarding the desirability of two or more alternative system configurations. For example, the recognition that graphical methods of information presentation exist in addition to the more familiar textual display naturally leads to a consideration of the merits of the two systems. Where possible, it has been indicated in this report how applied human factors methods may be used (and have been used) to resolve such issues. To the extent possible, the literature has been distilled in order to provide preliminary guidelines.

Lastly, in recognition of the need for multi-faceted design guidance, this report contains material that is not directly related to systems evaluation. There is a chapter of recommendations which are designed to complement the management practices of the procurer during a development contract. Recommendations for further research are also made. Appendix B discusses some observations made about the design process, and about the impact of human factors on that process.

(a)	Human-machine allocation of function
(b)	Use of decision-aiding systems
(c)	Hypothesis and option generation within C^2
(d)	Decision-making vs. stereotyped behaviour
(e)	Effects of operator strategies on system performance
(f)	Individual difference in performance
(g)	Effects of unreliable data
(h)	System flexibility
(i)	Manual back-up possibilities
(j)	Voice vs. non-voice communication
(k)	Graphical vs.textual displays
(l)	Team structure
(m)	Training
(n)	Personnel estimates

TABLE 1 - Some human factors issues in C^2 system development

2. DIAGRAMS TO DESCRIBE THE HUMAN AS A SYSTEM COMPONENT

2.1 Introduction: Uses and Definition

Diagrams are frequently used throughout the system development process. Human factors engineering employs diagramming techniques to clarify details of the human performance required and the interactions between the human and other system elements.

Diagrams have a number of general uses. They summarize chosen aspects of system functioning in a graphic form, that facilitates comprehension. In particular, they can make explicit the sequence of procedures in the system. Within a design context, possibly the most important function of diagrams is that they allow predictions of human performance to be made, and thus form the basis of systems evaluation from a human factors perspective. These predictions may be purely descriptive or, alternatively, may be quantitative if suitable data are inserted.

Diagrams may therefore function as preliminary models of the system. In fact, both diagrams and quantitative modelling techniques are frequently used together, i.e. the diagram provides a descriptive framework which is then translated into a formal computer model. Both techniques, however, constitute an abstract representation of system functioning. The major distinction is that diagrams provide a graphic description of human performance whereas modelling is most frequently associated with a computer simulation.

The diagram techniques which will be discussed in this chapter are presented in Table 2. It should be noted that the various diagrams tend to have different purposes, i.e. they represent different aspects of system functioning. For example, a broad distinction may be made between those diagrams that are based on a temporal sequence of activities (a,b,c) and those that concentrate on the information flow within the system (d,e,g,h,i).

TYPES OF DIAGRAMS

(a)	Functional flow diagrams
(b)	Spatial/temporal diagrams
(c)	Activity diagrams
(d)	Informationflow diagrams
(e)	Operational sequence diagrams
(f)	Network diagrams
(g)	HIPO charts
(h)	Job process charts
(i)	Process control diagrams

TABLE 2 - Types of diagrams incorporating the human as a system component

Correspondingly, a distinction may be made between those diagrams of overt human behaviour and those that attempt to model the human's cognitive processes. Some diagrams, e.g. operational sequence diagrams, can capture both aspects. However, in general terms, the order of presentation in this report is based on the distinction between activity sequences and information flow techniques. Given that a number of categories of diagram exist, representative examples of each are provided. It is hoped that related forms of diagrams not reviewed here may be accommodated within this classification.

Some critical evaluation also accompanies the diagrams. Briefly, the uses of each diagram technique are indicated and when it could be applied within the system development cycle. A discussion of criteria for the evaluation of models, that appears in a later section of this report (Section 3.2), is generally relevant to an evaluation of diagrams and should be read for further details.

2.2 Functional Flow Diagrams

Following the determination of system requirements, functional specification is a prerequisite for system design (Woodson, 1981). System functions may not necessarily be represented through a flow diagram; however, this technique certainly facilitates the consideration of alternative designs and trade-offs. A functional diagram acts as a gross, qualitative model of the system, before the functions of people and machines have been distinguished. The form of this preliminary diagram is dictated by an analysis of system goals and requirements. That is, although it is possible that different functions may achieve the system goals, mission analysis often reveals an optimum (or most efficient) set of functions (See Figure 1).

For example, Lindquist, Jones and Wingert (1971b), in the design of controls and displays for a search-and-rescue helicopter, derived system requirements through interviews with operators and through an analysis of similar systems. Typical mission scenarios were written down and then represented on an 'event flow diagram'. The presumption underlying this techique was that a typical mission would involve an orderly sequence of events, each with a predictable duration. Eight basic functions were then identified. Similarly, in an evaluation of naval bridge designs, Mara (1968) utilised a 'system flow diagram' which represented the functions involved in normal operation.

Apart from their illustrative value, the main use of functional flow diagrams is to assist in making decisions regarding the allocation of function between humans and machines. Given a preliminary functional diagram, it is then proper to ask which functions should best be automated. The basis of this decision may be a formal comparison of the relative abilities of the person versus the machine, e.g. Fitts (1951). However, the 'static' nature of this comparison has been criticised (Jordan, 1963). More commonly, an initial allocation is made subject to the provision that a re-allocation of functions may be necessary if later forecasts of workload for system personnel prove to be too great. Design thus proceeds in an iterative fashion, e.g. Mara (1968), Lindquist et al (1971c).

LEVELS OF FUNCTIONAL BLOCK DIAGRAMMING

LEVEL 1

1.0	2.0	3.0	4.0
DEVELOPMENT	PROCUREMENT	TESTING AND ACCEPTANCE	INSTALLATION AND CHECKOUT

OR

5.0 COMBAT INFORMATION SERVICE

6.0 MAINTENANCE

LEVEL 2

5.1	5.2	5.3	5.4
DETECTION AND TRACKING	IDENTIFICATION	THREAT ASSESSMENT	WEAPON ASSIGNMENT AND CONTROL

LEVEL 3

5.1.1	5.1.2	5.1.3	5.1.4
TENTATIVE TRACK DESIGNATION	TRACK CONFIRMATION	POSITION UPDATING	COURSE AND SPEED COMPUTATION

FIGURE 1 - Functional diagramming at three levels (Woodson, 1981)

System functions may also be represented along a time-line, that adds descriptive and predictive power to the diagrams. More specifically, these temporal data (if available) may influence decisions on allocation of function; for example, a common situation is that the time constraints for the performance of a function may be such that automation is considered necessary (Lindquist et al, 1971b). At a more detailed level, diagramming may assist the allocation of function between personnel. Typically, anticipated workload problems have been solved by a re-distribution of tasks amongst crew members, or by procedural changes (Mara, 1968; Lindquist et al, 1971c).

2.3 Spatial/Temporal Diagrams

Technically, functions of the system refer to relatively molar events such as detection, tracking, etc., which often do not distinguish between human and machine (Singleton, 1974). In practice, the terms 'functions', 'events', 'tasks' and 'activit:es' tend to be used interchangeably. That is, there is little common agreement between workers regarding systems description at different levels of abstraction. For the present, techniques which represent temporal or spatial aspects of the system will be discussed within the one context. An example is shown in Figure 2.

As regards temporally-based diagrams, it is possible to represent a number of aspects of a system along a time-line. For example, Lindquist et al (1971b) depicted the expected time course of system functions (for a search and rescue helicopter) before human and machine had been differentiated. Later in the design process, hypothetical time-lines were drawn for the activities of individual crew members. Combined with data concerning the temporal constraints that the system mission was expected to place on these activities, the diagrams facilitated an approximate form of workload estimation. More specifically, if performance time was predicted to be greater than that which was available, re-design was necessary (see Figure 3).

TIME LINE SHEET		FUNCTION: Propellant Transfer and Loading						
		Function/Tasks						
			Time (Hours)					
			0	1	2	3	4	5
TRANSFER	1.0	Connect Propellant Transporter to Fixed-Base Transporter System						
	2.0	Transfer Propellant from Transporter to RSV's						
	3.0	Drain/Purge/Decontaminate Lines, Equipment and Hardstand						
	4.0	Condition Fuel						
	5.0	Drain/Purge/Decontaminate Lines, Equipment and Conditioning Unit						
LOADING	6.0	Perform SLV Tank Purge/Leak Check						
	7.0	Connect Propellant Transfer Umbilicals						
	8.0	Transfer Propellant from RSV to SLV-Stage I						
	9.0	Monitor Propellant Transfer						
	10.0	Transfer Propellant from RSV to SLV-Stage II						
	11.0	Monitor Propellant Transfer						
	12.0	Drain/Purge/Decontaminate Propellant System						
	13.0	Secure System						

FIGURE 2- System functions for a propellent loading task, represented along a time line (Woodson, 1981)

FIGURE 3 - **Time line for the activities of five operators. Workload was defined as the ratio of time required to time available (Lindquist et al, 1971b)**

In a similar fashion, Baker (1970) developed a diagram of the average time course of data flow through the U.S. Army's current Tactical Operations System. At a more detailed level, individual tasks were identified and given a temporal representation. Estimation of performance time and constraints was used to make predictions regarding possible information 'bottlenecks' in the system. Estimation of human error rates was also attempted from the activity time-line. That is, a presumption was made that tasks which were subject to relatively great time constraints would be error-prone. Pew, Woods, Stevens and Weene (1978), also analysed a military information system (part of the Tactical Air Control Center) by representing selected tasks along a time-line. This technique was the first step towards developing a procedural description of activities within the system.

With regard to spatial representations of systems, diversity in the type and level of detail of the diagramming also exists. For example, Lindquist et al (1971b) constructed 'profile descriptions' for the typical missions of a proposed search and rescue helicopter in order that travelling distances and altitudes could be represented. These diagrams assisted the derivation of system requirements. The most common use of spatial diagramming, however, is to assist the evaluation of workspace layouts; particularly if the technique of link analysis is applied - (see Section 2.4).

2.4 Activity Diagrams

At conceptual stages of design, spatial/temporal diagrams are concerned with molar units of behaviour, namely, system functions. As the development cycle proceeds, however, the level of detail which it is

possible to represent increases. Typically, the design of controls, displays and workspace layouts becomes an issue relatively late in the development cycle. Accordingly, it becomes necessary to predict (and represent) the actions of individual operators in some detail.

Historically, methods engineers were the first to make widespread use of task diagramming in their investigations of the activities that industrial workers perform (McCormick, 1979). The tasks were represented as a sequence of coded activities, with the main application being attempts to devise the most economical sequence of motions. Later, time study also became an area of concern, so the motion diagrams were represented along a time-line. The human factors community has embraced these techniques under the title of 'activity analysis' suggesting that the diagrams serve a greater purpose than the mere representation of data. In fact, activity analysis diagrams often facilitate the identification of task-related difficulties of the operator. For example, it may be that the movements of the operator are inefficient, or even incompatible, or there may be unrealistic time constraints for the execution of some movements. That is, the activity diagrams allow one form of analysis and evaluation, albeit in a rather subjective fashion (see Figure 4).

A related technique is link analysis (Chapanis, 1962). Links are drawn between the elements of a system (e.g. between people or between people and machines) in order to represent the frequency of contact, or communication, between them. The diagrams therefore allow representation of statistical data. With the presumption that frequent communicators within a system should be easily accessible to each other, then link analysis may be used to solve workspace layout problems. The technique has been used for the re-design of a naval command station (Chapanis, 1962) and could be applied to the evaluation of a conceptual design. See Figure 5 for examples.

Christensen (1971) has made the criticism that the users of time-and-motion based techniques have tended to neglect individual differences in performance. In principle, this limitation may be overcome. However, time-and-motion based diagrams tend to represent system aspects from a limited psychological viewpoint. The relevant operator behaviours are usually represented by clearly defined outputs and they generally occur in a fixed sequence during any one task. In C^2 systems, it is clear that analysis of more than manually repetitive tasks is required. Monitoring and decision-making are important functions, and a comprehensive task analysis must therefore take into account the information requirements of the operator. These aspects of operator behaviour are more abstract than is conventionally recorded in a time-and-motion analysis.

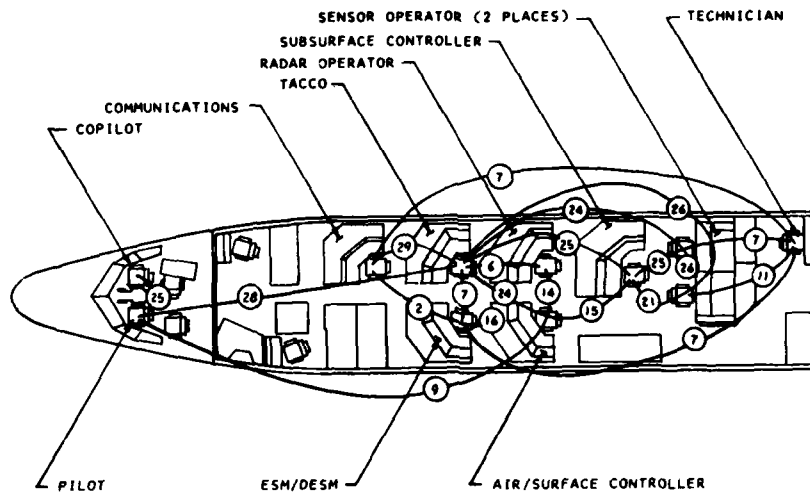## 2.5 Information Flow Diagrams

The information flow through a system, or through a task, has become the major paradigm by which a systems psychological phenomena are investigated (De Greene, 1970). This orientation is also reflected in the high priority placed on communications analysis in modern systems. Various forms of information flow diagram exist in order to represent this factor.

FIGURE 4 - Activity analysis of a pilot during aircraft landing (Chapanis, 1962)

NUMBER IN CIRCLE PRESENTS RELATIVE CRITICALITY X FREQUENCY
OF COMMUNICATIONS FOR A GIVEN FUNCTION/TIME PERIOD

Original Panel   Modified Panel
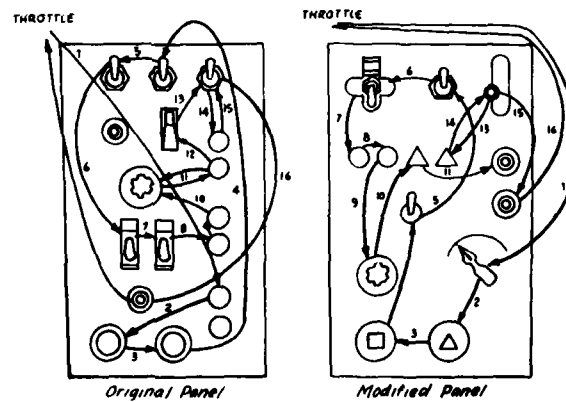
FIGURE 5 - Two examples of link analysis (Woodson, 1981)

Once again, it is possible to use a diagram to represent informational aspects of the system at different levels of abstraction. At a macro level, for example, one may plot the information channels of a production system (Figure 6). More commonly, it is necessary to represent the information flow through the system by means of a branching chart. Such a chart depicts the conditions upon which actions in the system depend, e.g. Figure 7. From a design perspective, these diagrams are most useful for assisting 'allocation of function' issues. In particular, the identification of situations of high information processing load may suggest that full or partial automation may be necessary (an example of the latter being a decision aid for the human). Alternatively, if there is a requirement for the integration of novel or unexpected events, the talents of the human operator may need to be highlighted.

At a psychological level, the functions of the operator may be distinguished from the functions of the machine (Figure 8). As operator actions may be seen to depend on the receipt of certain information the role of human decision-making needs to be made explicit. Branching charts are a particularly suitable means of representing information flow as they facilitate generation of a computer algorithm in the next step towards modelling the system. These diagrams may most easily be constructed when the mission is sufficiently well-defined so that an optimal sequence of actions can be specified. On the other hand, the diagrams may not provide a good description of an individual's behaviour. This may apply in many process control operations (Drury, 1976).
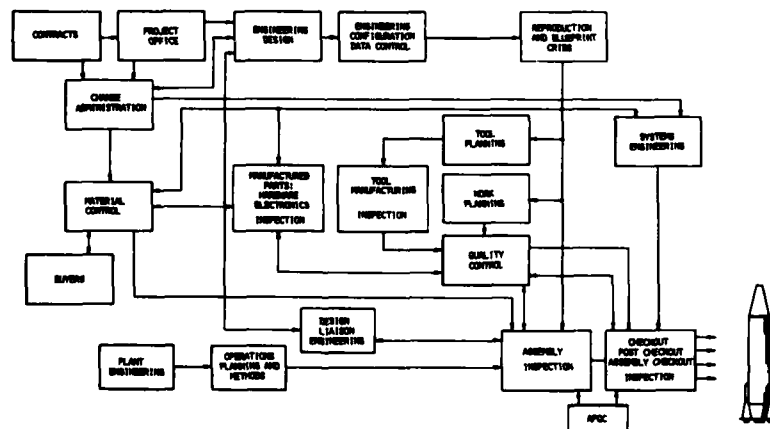
FIGURE 6 – Information flow within a production system (Meister and Rabideau, 1965)

Information-based diagrams help to overcome some of the limitations of activity analysis; namely, the difficulties with representing cognitive behaviour. Information flow diagrams are often considered to be a supplement to activity analysis. For example, in an analysis of a U.S. Air Force control centre, Pew et al (1978) used both kinds of diagramming technique in order to make recommendations about which functions should be computerised in future. The activity diagrams facilitated identification of undesirable time constraints for task performance, whilst 'algorithmic flow' diagrams contributed to the identification of points of both information overload and monotony.

Example: Gross-Level Flow Chart for
Detection and Tracking

Start

Monitor incoming signals from
  Surveillance System

Compare signals with previous
  target list

N   Any new, probable targets?
  Y

Enter tentatively into system
  memory

N   Does probable target reappear?
  Y

Drop tentative from system
  memory

Confirm as target in system
  memory

Generate initial course/speed
  from elapsed time/displacement

Update all target positions as
  necessary for tracking

N   Any target signals disappear for
  Y   critical time?

Drop target from system memory

Note that no human-machine function alloca-
tion has been assumed at this level

FIGURE 7 - **Information flow chart for a hypothetical detection and tracking system (Woodson, 1981)**

Design applications of information flow diagrams appear to be few in number in the literature. This lack probably stems from the relative difficulty of forecasting cognitive requirements (from a scenario) at preliminary phases of development. Galer (1979), for example, emphasised

this difficulty whilst discussing the role of human factors within transport system development.

Start

Any target tracks in system?

Press SEQ button

Put next target in track list under close control

Advance hook on CRT to coordinates for track under close control

Is target video present?

Does hook line up with present target position?

Enable track ball/reposition to move hook over target

Press POS CORR button

Add latest position data together with time to to memory. Compute/store course/speed. Periodically update target position

Any target fail to be updated within critical time?

Display "Recommended Drop Track" alert

Drop alerted track?

Hook/press DROP TRACK button

Delete track from memory

○ Human Operations   ● Machine Op

◇ Human Decisions   ◆ Machine Dec

FIGURE 8 - **Information flow chart showing a distinction between operator and machine functions (Woodson, 1981)**

2.6 Operational Sequence Diagrams

One technique which includes all of the diagramming features discussed so far is the operational sequence diagram. It may be regarded as an information flow diagram set against a time-line of coded activities (see Figure 9). Kurke (1961) also claims that a spatial analysis of movement or communication within a system may be represented on an operational sequence diagram, although this does not appear to be a primary use.

These diagrams are also claimed by Kurke to have a specific conceptual design application. That is, the operational sequence diagram

A TRADITIONAL METHOD   B PROPOSED SYSTEM
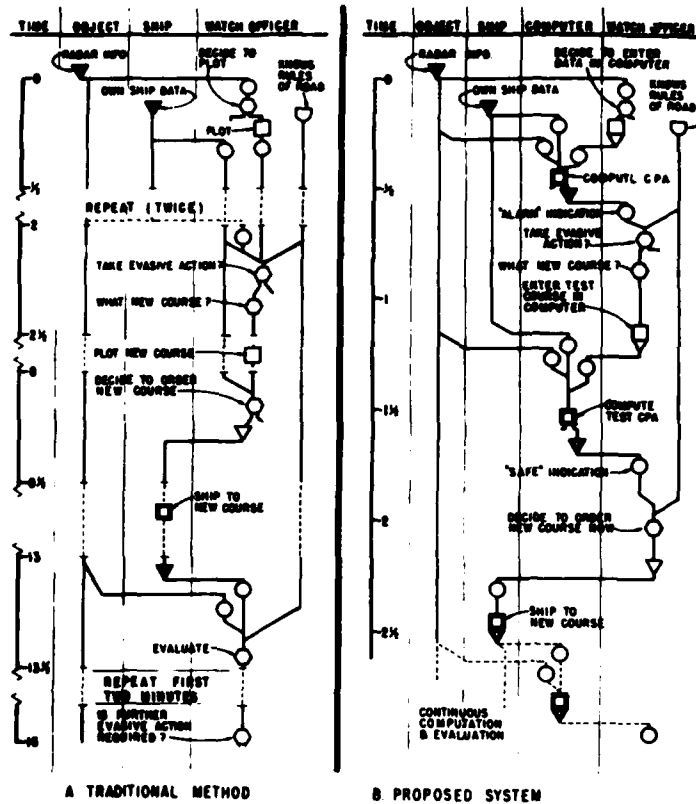
FIGURE 9 - Operational sequence diagrams for the manual ship
            avoidance system (on left) and a proposed semi-
            automated system (Kurke, 1961)

may be analysed into a logical network and may thence be given a symbolic representation (see Figure 10). The reliabilities of the various logical sequences may be calculated in order to evaluate the desirability of alternative systems. As discussed in the section on reliability modelling in this report, the success of such forecasts depends, amongst other things, upon the availability of a human reliability data base. Network diagrams are also discussed more fully in the next section.

Kurke (1961) proposed an operational sequence diagram for a ship avoidance system; however, the system was only hypothetical. Mara (1968) claimed to have used these diagrams in evaluating alternative naval bridge designs, however, no details were given. It is therefore difficult to gauge the efficency of the technique. The use of operational sequence diagrams is probably more widespread than reports in the open literature may suggest. Within the U.S. Naval Air Development Center, for example, the diagrams are regarded as a standard technique (Parks & Springer, 1976).
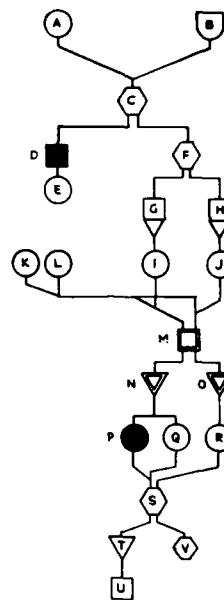
Analysis of
alternate actions

$A_1 \cdot B_1 \to C_1 \to [(D_1 \cdot F_0) \to E_1] + (D_0 F_1)$

$F_1 \to [(G_1 \to I_1) \cdot H_0] + [G_0 \cdot (H_1 \cdot J_1)] \cdot \dot{M}$

$K_1 \cdot L_1 \cdot (I_1 + J_1) \to M_1$

$(H_1 \to J_1) \cdot K_1 \cdot L_1 \to M_0 \to (N_1 \cdot O_0)$

$\to (P_0 Q_1) \cdot S_0 \to T_1 \to U_1$

$(G_1 \to I_1) \cdot K_1 \to L_1 \to M_0 \to (N_0 \cdot O_1) \to R_1 \to S_0 \to V_1$

FIGURE 10 - Logical network of a proposed navigation system (Kurke, 1961)

2.7 Network Diagrams

It has frequently been implied throughout this review that system diagrams may function as qualitative models of performance. Hence, the diagrams permit a crude form of performance forecasting, and these predictions may be enhanced by the use of a quantitative data base.

Network diagrams are based on the 'bottom-up' approach (Pew et al 1977) to systems modelling. With this approach, the system is first analysed into discrete 'events'. As shown in the network example of Kurke (1961) in Figure 9, these events may include both decisions and actions, and may be performed by either human or machine. Each event then forms a node in a network tree. That is, there may be a number of possible outcomes at each node. Tracing along one branch of the tree represents a particular event sequence. If suitable data exist (namely, reliabilities or performance times for each event), the event sequence may then be described quantitatively.

Within systems, it is rare that a fixed sequence of events will always occur. Network diagrams may incorporate all possible outcomes, and thus allow a comprehensive systems representation. Given a quantitative prediction for each event sequence, that sequence may also be weighted by its probability of occurrence in order to yield an overall systems evaluation. This estimate may be obtained analytically, i.e. by some mathematical equation, or via simulation. In either case, it could be said that the network diagram forms the basis of a system's model.

A related, but inverse technique is fault-tree analysis (De Greene, 1970). Commencing with the identification of an error, an attempt is made to trace the possible events that led to that error. A system's network may then be constructed in the standard manner (e.g. Figure 11). The level of detail of this network may also be increased in order to include cognitive behaviour. An example from the area of nuclear process control is provided by the 'Murphy diagrams' of Pew, Miller and Feehrer, 1981 (see Figure 12). The use of fault-trees as described here is more concerned with the retrospective analysis of operational systems than with evaluation of conceptual designs (i.e. prediction). The diagrams are an aid to visualization and, in the study by Pew et al (1981), the technique was used to promote discussion amongst an expert panel regarding crucial failure nodes in a nuclear system.
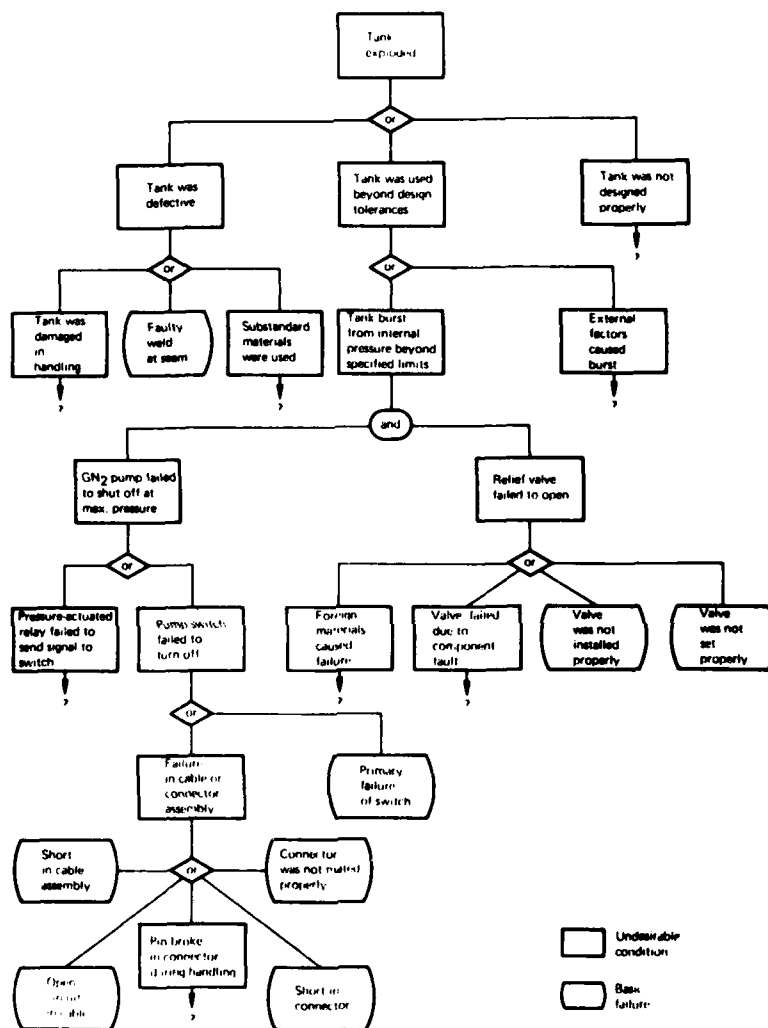
FIGURE 11 -   Fault tree analysis of the possible events leading
               to a tank explosion (Cornell, 1968)
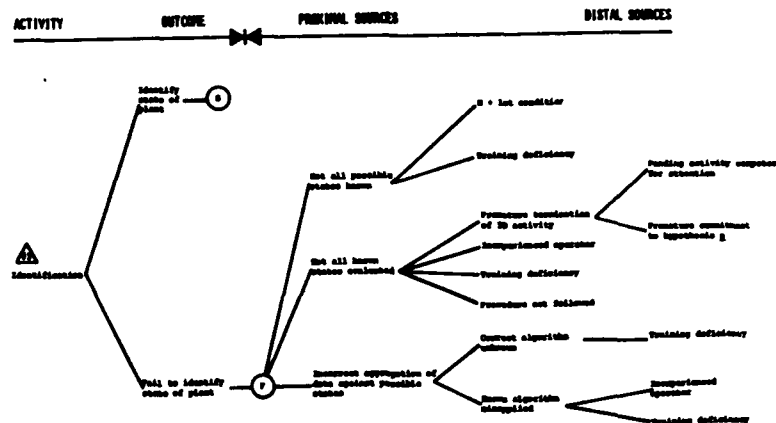
FIGURE 12 - 'Murphy' diagram for identification of nuclear plant
state (Pew, Miller and Feehrer, 1981)

Recently, diagrams that have a more specific modelling orientation have been developed. These are part of the Structured Analysis and Design Technique: SADT (Davis, 1982) and the Integrated Computer Aided Manufacturing Definition Language: IDEF (Wohl, Entin, Alexandridis & Eterno, 1983). Both these techniques are said to assist the modelling of human performance in systems by structuring the manner in which the system is first analysed, or decomposed. That is, system goals are used to specify the level of detail at which analysis and subsequent diagramming are best performed (see Figure 13). This 'top-down' approach may avoid a common problem in modelling, namely that of choosing irrelevant operator behaviours (Davis, 1982). Once the system has been decomposed, performance forecasts are made by the use of a network simulation language, usually SAINT (Pritsker, Wortman, Seum, Chubb & Seifert, 1974). (For a more comprehensive treatment of this topic, the section on modelling in this report should be consulted as well). Wohl et al (1983) claim that a practical benefit of IDEF diagrams is that they may reduce time spent on SAINT modelling by 60%.

IDEF diagrams have also been developed in response to the need for better representation of human decision-making in systems. These diagrams basically represent the information flow in the system. Activities which are conditional upon the receipt of certain information may then be modelled. Probabilistic relationships may be incorporated. A further refinement is that the single decision itself may be analysed and then represented in a standardised fashion, via the stimulus-hypothesis-options-response (SHOR) paradigm of decision making (Wohl, 1981). This technique allows the diagramming of more detailed factors which contribute to a decision. In particular, it is frequently the case that actions are not simply contingent upon the receipt of certain stimuli; rather, the
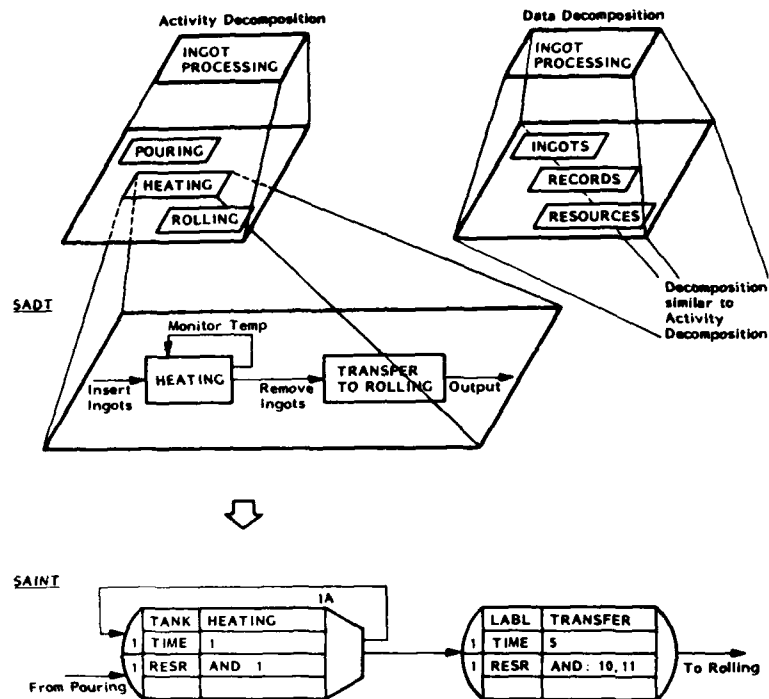
FIGURE 13 -   **An elxample of the decomposition of an ingot process using SADT (Davis, 1982)**

decision maker must also form hypotheses about the state of the world and generate the options which are available. IDEF diagrams may incorporate these factors, and thus promote the modelling of the effects of both stimulus uncertainty and consequence-of-action uncertainty on human performance.

2.8 HIPO Charts

Hierarchical input-process-output (HIPO) charts appear to be a generic technique within computer system design (Brookes et al, 1982), although few details of their use have been found. This type of chart is included in this review because of one study (Montgomery, Thompson & Katter, 1980) which had a direct human factors application.

Basically, the charts depict human activity in a hierarchical fashion, through a flow diagram. Three levels of activity have been distinguished: activities, processes and sub-processes. That is, any one activity may consist of a number of processes and a further number of sub-

processes for each process. Simultaneously, the inputs which motivate each activity/process may be represented, along with the resulting outputs. The charts thus conform to a stimulus-organism-response paradigm of performance.

Montgomery et al (1980) were concerned with the formulation of a model of the processes underlying intelligence analysis (both strategic and tactical). HIPO charts were selected because they allow a representation of behaviour which is not necessarily sequential in nature and which is covert. The charts were constructed after data had been collected through interviews, observation and document review. An example is shown in Table 3. The major goal of the construction of the intelligence model is to assist in the reorganization of part of the U.S. Army's intelligence system, including design modifications such as further automation (through decision-aiding) and training modifications such as procedural change.

Despite the claims of the authors, evidence is not yet available to confirm that the charts promote an adequate representation of cognitive behaviour. The units of behaviour shown in Table 3 are relatively molar, although there is no reasons in principle why some of the sub-processes could not be further analysed. A further criticism is that it is uncertain whether the charts can accommodate quantitative data and thus facilitate performance prediction in addition to performance description. At present, it appears that the charts must be derived from an operational system. Within a design context, the charts appear to be more useful for system re-design rather than for conceptual design.

INPUTS	SUBPROCESSES	OUTPUTS
Imagery Debriefing/mission reports	Identify indications of energy camouflage and concealment activities.	
Maps/Overlays	Identify possible dummy positions, as well as	
Equipment keys	possible dummy equipment and decoys.	
Reference material	Identify abstracts, barriers, choke points, fortifications and other defences.	Identification data
	Identify possible enemy supply areas.	
	Identify bivouacs, head-quarters units, and other installations.	
	Identify associated personnel	

TABLE 3 — HIPO chart for the process of identification of items of military significance (Montogomery et al, 1980)

2.9 Job Process Charts

Job process charts (Tainsh, 1982) have been developed specifically for representing human-computer interactions in a tactical situation. In many ways, they represent a standard information flow chart, i.e. they represent information-decision-action sequences. It is possible to formulate the charts at different levels of abstraction so, for example, it may not be necessary to distinguish the activities of human and machine. More commonly, it is necessary to distinguish the activities of the operator, the machine and also the contents of their transactions, in separate columns (see Figure 14). It may be seen that the transfer of information between human and computer is delineated.

No design applications of job process charts have been found in the literature, although it is possible to make some speculative comments. Tainsh (1982) believes that the charts may be useful for assessing a number of software-related issues, e.g. the organisation of displays and the amount of information on each. The identification of informational cues that are important to the user may suggest decision-aiding or training requirements (Tainsh, 1983). Given a suitable data base, it is possible to estimate task times and error rates.

The applicability of job process charts at conceptual stages of design may be low, due to the difficulty of anticipating operator-computer transactions in detail. Given an operational system, however, re-design may be suggested. For example, Tainsh (1983) made a hypothetical contrast between a graphical and an alphanumeric-based task.

2.10 Process Control Diagrams

Process control diagrams have been developed within an industrial setting, but have relevance to the supervisory control tasks of C^2 systems. The method of diagramming is based on the signal flow graph (Beishon, 1967), in which the relationship between system variables is described rather than that between physical entities (see Figure 15). The technique may be used for identifying a process operator's control strategies (Drury, 1983), i.e. it allows a representation of an operator's perception of the relationship between system variables, and gives a qualitative indication of the control actions that are necessary (see Figure 16).

Process control diagrams may be derived by two different methods. One relies on a logical analysis of the process to be controlled, and yields a prescriptive control model (Bainbridge, Beishon, Hemming and Splaine, 1968). The second is basically descriptive in construct, and relies on an analysis of operator actions and/or protocols (Bainbridge, et al, 1968; Rasmussen, 1980). In practice, it is often difficult to formulate a prescriptive model and the descriptive model is often obtained in the absence of a 'true' model of the process (Drury, 1976). Any descriptive model may then be compared with that obtained from expert controllers in order to assess its normative content, or in other words its degree of validity.

No examples of a design application of process control diagrams to armed services projects could be located. Logically, the relevance of the technique at conceptual stages of design is probably low, due to the
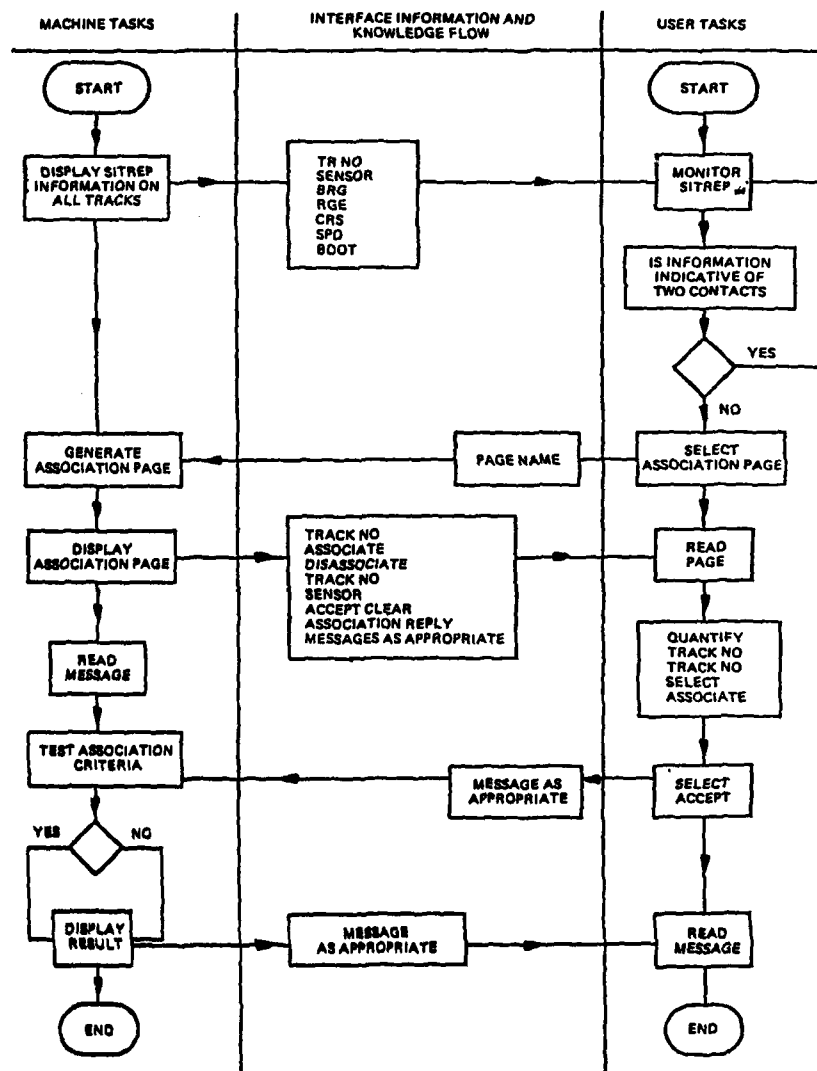
FIGURE 14 - Job process chart for track association during a
tactical 'target motion analysis' task (Tainsh, 1983)
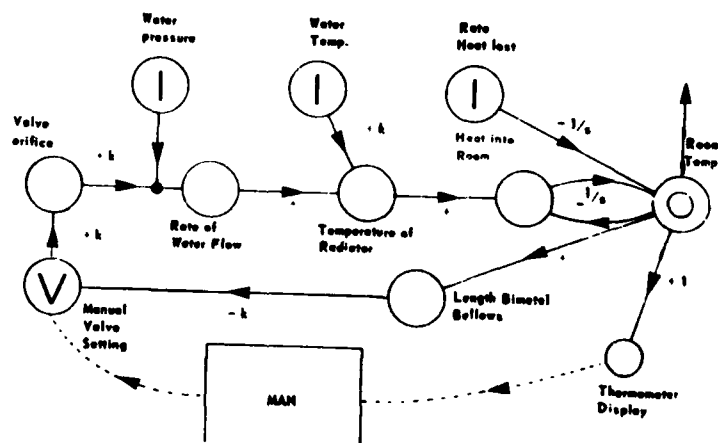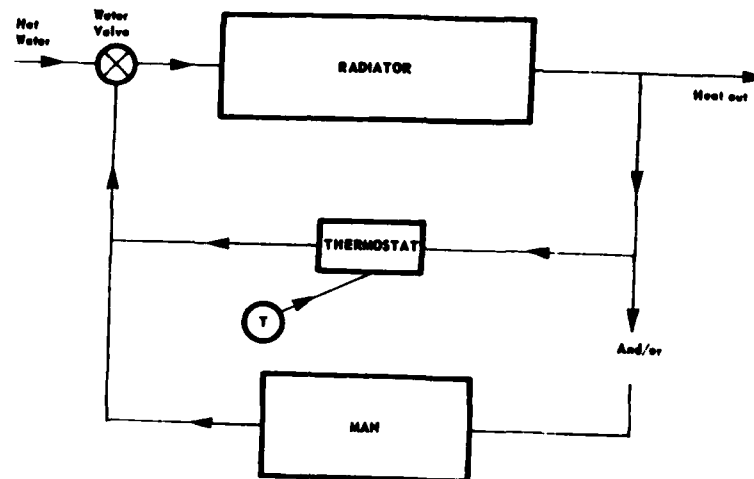
FIGURE 15 - Block diagram (top) and signal-flow graph for control
of room temperature (Beishon, 1967)
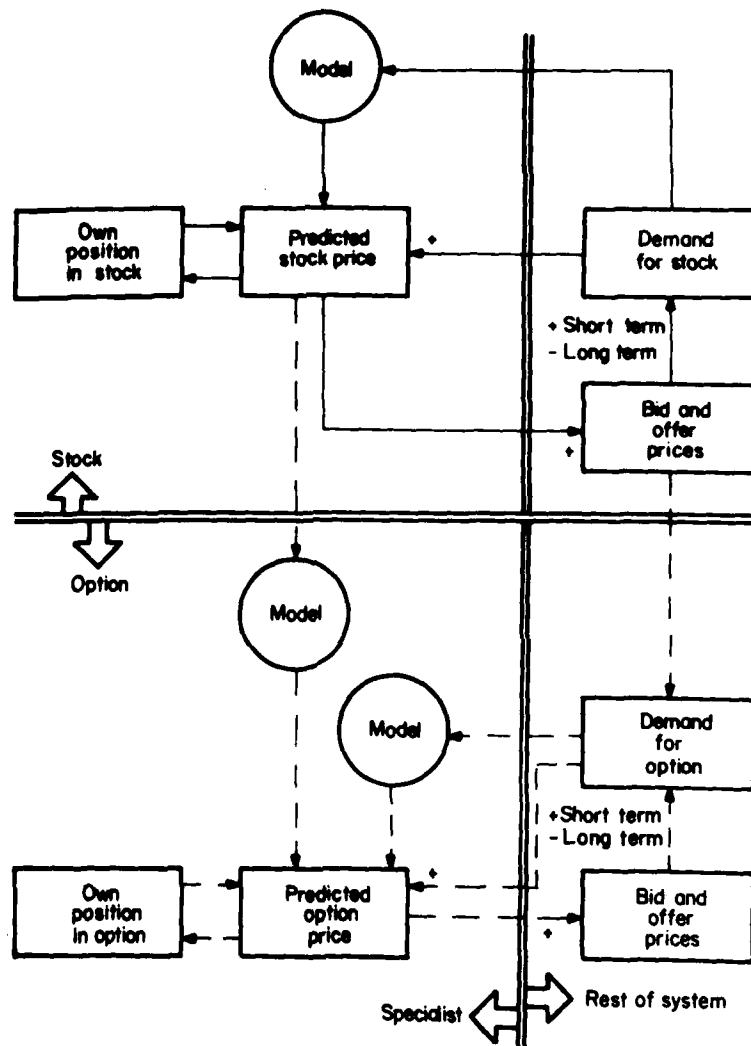
FIGURE 16 - Task description of stock and options trading process
through the signal-flow graph method (Drury, 1983)

difficulty of formulating a process control diagram in the absence of an operational system. A more realistic goal may be that the diagrams can assist system re-design through the identification of a process operator's control strategies. That is, the diagrams provide a focus on the relevant system variables and thus may indirectly suggest a need for improved informational or control capability. Given a prescriptive diagram, training programs may be suggested. That is, the diagrams promote a specialized analysis of certain tasks, that is a prerequisite for training programme development (Royal Australian Navy (RAN) School of Training Technology, 1978).

2.11 Summary

The discussion of human performance diagramming has taken place within a framework utilising two evaluation criteria; namely, the purpose diagrams serve and, secondly, the relevance of diagrams at various stages of design. We have tended to make a distinction between those diagrams that either capture cognitive behaviour or informational aspects of the system, and those that do not (although, in practice, all methods differ from each other widely on this dimension).

We have also distinguished between those techniques that are applicable at preliminary stages of design, and those that are not. The reason for both these distinctions arises from our oft-repeated concern for human factors in computer systems. We believe that, to be effective, human factors input to those systems should commence at an early phase of design and should address the design of the system as it relates to the cognitive performance of the operator. In other words, it is important to give attention to the 'cognitive engineering' of the system.

Table 4 summarises the conclusions which may be drawn from this review. The table illustrates whether or not each method has satisfied, or may potentially satisfy, the two evaluative criteria. (It should be noted that we are uncertain about the status of network diagrams). A general conclusion is that few diagrams are both applicable at preliminary phases of design and to the representation of cognitive behaviour. This conclusion was not unexpected, given the difficulty of forecasting cognitive behaviour from an unembellished design. In practice, multiple techniques are used to achieve human factors input. That is, relatively crude drawings and models are applied at early phases of design, and these techniques become more refined (and may address behaviour which is more cognitive) as design proceeds.

CRITERIA

	Early Design Application	Presentation of Cognitive Behaviour

TYPES OF DIAGRAMS

		Early Design Application	Presentation of Cognitive Behaviour
(a)	Functional Flow diagrams	Yes	No
(b)	Spatial/temporal diagrams	Yes	No
(c)	Activity diagrams	No	No
(d)	Information flow diagrams	Yes	Yes
(e)	Operational sequence diagrams	Yes	Yes
(f)	Network diagrams	?	?
(g)	HIPO charts	No	Yes
(h)	Job process charts	No	Yes
(i)	Process control diagrams	No	Yes

TABLE 4 - Applicability of human performance diagrams to two criteria

3. THE ROLE OF HUMAN PERFORMANCE MODELS

3.1 Introduction

Models occupy an important position in systems development. Hardware engineers commonly build prototypes of their design (and subject them to operational conditions) in order to gain an understanding of how the system is likely to function. Alternatively, the mock-up may be dispensed with altogether if it is possible to model the working system via a computer program. Modelling provides a check on the adequacy of a particular hardware design, and there is also a possibility of testing alternative designs. In a similar fashion, human factors engineers may utilise models as a means of forecasting the effects of a proposed system upon human performance.

At the most general level, models may be defined as analogies of the human (Chapanis, 1961). This immediately leads to a philosophical problem, namely, whether a distinction should be made between a model and a theory of human performance. In particular, it may be claimed that a theory should have some degree of explanatory power, whereas a model may be purely descriptive. On the other hand, this issue is probably not significant within the field of human engineering, as models are primarily assessed by their usefulness, for example, in making predictions (Pew and Baron, 1982). Modelling may in one sense be regarded as a very abstract means of 'collecting' information concerning performance, in contrast to a formal experimental evaluation (Obermayer, 1964). A large difference, of course, is that models forecast performance whereas performance is observed directly during experimentation or prototype testing.

3.1.1 Modelling vs. prototypes/mock-ups

Modelling of human performance should be distinguished from the situation in which people interact with a prototype, or mock-up, as a means of systems evaluation. While both types of techniques legitimately belong to the category of 'simulation', human behaviour itself is simulated during modelling, usually via a computer program.

The most significant advantage of modelling is that it allows a performance test of designs before those designs have been realised in hardware or software. Human engineering problems may thus be identified and costly, re-designs may be avoided. Hardware simulators are also used to provide a design check, but it is a frequent practice to build prototypes only after the design configuration has been settled (Price, Fiorello, Lowry, Smith and Kidd, 1980). Modelling thus permits a relatively early human factors input, at a stage before some of the prevasive design decisions have been made. Human performance modelling may be seen as an 'intermediate' aid to system design (Baron, Feehrer, Muralidharan, Pew and Horowitz, 1982). The technique often precedes the building of prototypes, but follows the initial descriptions and conceptual drawings of system functioning.

Siegel, Leahy and Wolf (1978) have also emphasised the flexibility of modelling. Naturally, the technique permits a form of systems evaluation without the necessity of collecting data from an operational system, or even from an operational prototype. Not only is the problem of recruiting suitable subjects for systems testing bypassed, but it may also be possible to simulate the activities of a number of

subject 'types' in a methodical fashion, i.e. modelling of individual
differences is possible. If a computerised simulation is chosen, as is
almost exclusively the case, modelling results may be obtained at a much
faster rate than if live subjects were used. Models may also be rapidly
modified in order to test the effects of a proposed re-configuration in
the design.

Modelling compares favourably with hardware simulation in terms
of reliability and dependability (Siegel et al, 1978). As an evaluative
technique, it is usually superior in terms of cost and time. Typically,
model development time is a significant cost, but this cost may be offset
if one is able to purchase a suitable computer package. The actual
computer running time is said to be a much smaller cost by comparison
(Siegel and Wolf, 1981).

3.1.2 Other uses of models

Human performance modelling also has uses which could broadly be
termed 'heuristic'. The act of formulating a model requires organisation
of performance issues (Rouse, 1980) and forces consideration of what might
otherwise be neglected aspects of the design problem (Pew and Baron,
1982). Related benefits are that models allow the visualisation of what
might be new performance relationships (Chapanis, 1961) and provide a
systematic framework around which to organise facts in a way that reduces
the memory load of the investigator (Pew and Baron, 1982). Even when
working with a pre-programmed model, the requirement for parameter
estimation may focus attention upon aspects of performance (and aspects of
design) that the modeller has previously considered to be important.
Models may be differentially sensitive to their parameters, so that
changes in the value of some parameters have a greater influence than
others on the model output. Such a sensitivity analysis may allow
anticipation of what are likely to be the important variables in later
prototype evaluations and experiments.

Lastly, modelling aids aspects of systems development other than
hardware/software design. Models permit a forecast of the procedures and
tasks which will occur in the operational system, so personnel-related
issues may be addressed at an early stage. The most useful models should
allow a forecast of the numbers and training levels of the personnel who
will be required to operate the system (Meister, 1971b). Alternatively,
the model may provide a check on the personnel-related specifications of
the development contract. Provided this latter goal has been achieved, it
should then be possible to devise job analyses, training manuals and
training programs in parallel with hardware/software development (Meister,
1971a).

3.2 Model Evaluation

In the next section of this report, the details of a number of
human performance models will be discussed. Logically, such a
presentation leads to the question of how models should be evaluated,
especially with regard to their application during system design. Models
may be evaluated along various dimensions which have greater or lesser
relevance to systems design, and such a critique will be attempted. In
this section, however, some modelling issues will be discussed in general
terms by way of preparation.

3.2.1 Validity

Probably the greatest concern when evaluating models is their validity, and the importance of this issue has been previously confirmed by a survey of human factors workers (Meister, 1971b). Broadly, the validity of a model refers to the extent to which it 'captures' the performance of interest. A common means of testing the validity of a model is therefore to compare the predictions of the model with actual human performance under similar conditions. In practice, however, tests of validity are not quite so straighforward, due to a number of interacting factors that contribute to validity.

Van Horn (1971) has distinguished three methods of testing the results of a simulation exercise. These are:

(a) Verification

(b) Validation

(c) Problem analysis.

Verification ensures that the model behaves as intended. This test largely reflects the ability to transfer abstract model concepts into a logical computer program. The test is independent of the collection of real-world data; all that is required is for the simulation program to be run a sufficient number of times so that a check on internal consistency may be made. Problem analysis refers to the ability of the model to focus on the performance of interest (and possibly to suggest solutions in a systems design context). Such a criterion is therefore pragmatic, as it is a large determinant of the 'usefulness' of the model.

Given that these two requirements have been satisfied, it is then necessary to attend to the validity of the model in a formal sense. Validity, depending on one's terminology, may once again be analyzed into three factors. These are:

(a) Reliability

(b) Construct validity

(c) Predictive validity.

A model is reliable if repeated applications under the same conditions yield similar results. This requirement does nothing to ensure that the model has actually captured the human performance of interest, although reliability is a necessary condition for validity. Analogously to a psychological test, a model may be reliable and invalid, but not vice versa. Construct validity refers to the extent to which the human processes represented by the model are similar to those which are thought to occur in reality. In the case of cognitive behaviour, this criterion is tested by comparing the model's representation of cognition with that which may be inferred from observable behaviour. Needless to say, no model has ideal construct validity, but different models may be ranked according to this criterion.

Predictive power is the essence of the popular notion of validity, and refers (as previously discussed) to the congruency of the model's predictions with human performance data that have been obtained empirically. A high degree of construct validity of a human performance

model frequently helps to ensure predictive validity, but this is not a necessary relationship. In fact, the optimal control model (Kleinman, Baron and Levison, 1970), has made the most successful forecasts under certain conditions whilst retaining many simple assumptions about human performance. On the other hand, it could be said that the most frequent strategy of model-builders when faced with a model of inadequate predictive power is to modify the constructs, usually in the direction of greater detail. This issue will be discussed more fully in the following section.

The act of predictive validation of a model invites a logical fallacy (Chapanis, 1961). That is, when executing the original simulation program, human performance consequences are deduced from a certain set of input data and a certain model structure. If those consequences are then observed in reality, it is still not possible to claim that the model's constructs accurately reflect human behavioural processes. Put differently, in principle the construct validity of a model can never be completely ensured. In a similar fashion to psychological theories, the truth value of models cannot be proved, only disproved. Construct validation should therefore be seen as a process of gaining evidence which increases confidence in the model according to one's purposes, i.e. confirmation is possible.

In a design context, the predictive power of human performance models is their most important feature. Designers need to be able to assess the adequacy of their concepts from a human engineering point of view without the necessity of testing live subjects or building prototypes. As discussed previously, the act of constructing a model may promote a number of insights, but these benefits should be regarded as secondary. A model's construct validity should be sufficient to ensure predictive validity to appropriate degree in the required circumstances, i.e. validity is highly situational (Lane, Strieb, Glenn and Wherry, 1980). In most cases, designers would probably wish to familiarize themselves with model constructs only insofar as is necessary to implement the model.

An inherent difficulty with testing model validities is the comparison between model predictions and empirical data. If this comparison is performed statistically (via the null hypothesis), a non-significant difference between model output and empirical data suggests that the model has predictive validity. This is the reverse strategy to that which is applied when attempting to test experimental hypotheses. Circumstances which increase the power of the statistical test (such as the use of larger data samples, less variable subjects or more sensitive tests) increase the probability that the model will be rejected as invalid. One solution to this problem is to use correlational statistics. However, a far more common procedure seems to be an informal comparison based on the subjective judgement of individuals or expert panels.

In one sense, the use of models is tautological (Chapanis, 1961). Experiments assist the process of gaining information about the world through hypothesis confirmation and rejection. In modelling, by contrast, the results are in a way pre-determined; model output is an exclusive function of input data and model structure. If model predictions are incongruent with empirical data, then it is usually the case that the model structure or parameter values are changed rather than

the conditions of the evaluation procedure being scrutinised more closely. As an example, prescriptive models of performance are not developed to predict idiosyncratic behaviour. However, if such behaviour is observed during an experimental validation and is regarded as essential, then idiosyncracies may be programmed into the model for the future.

The method of devising a model for prediction of human performance (as an aid to design, for example) is a three-stage process. The model is first constructed from a set of observations. It is then validated under conditions which are both similar to and different from the original conditions. The new conditions must be sufficiently similar in order for the model to be applicable, i.e., no model of human performance is all-inclusive. On the other hand, if the new conditions are identical to the original conditions, then the model's generality has not been established and its predictive power is compromised (Silvern, 1970). For this latter reason, instances in which the model is validated against the same data set that it is supposed to predict are methodologically unsound (Miller et al, 1978). In such circumstances, what passes for validation may merely involve adjustment of model structure and parameters until an adequate fit to experimental data is found. Pew, Baron, Feehrer and Miller (1977) prefer to label such an undertaking as model 'identification'. Unfortunately, it is a necessary process when the values of model parameters cannot be specified (either through theory or past experience) in advance of empirical data.

The third stage of modeling involves simulation of the conditions implied by some conceptual design, in which the model outputs are used to assess the consequences of the design. The accuracy of the model's predictions based on a design can only be tested in retrospect, i.e., by observing the operational system or a prototype. Paradoxically, if circumstances exist which ensure the predictive validity of a model to a high degree of precision, than the model becomes superfluous under those conditions in favour of direct experimental data (Obermayer, 1964; Pew et al, 1977). In the interim, the best strategy is to observe the accuracy of the model in predicting the behaviour of similar systems. If this cannot be achieved, a common alternative is to check the sensitivity of the model to certain parameters against prior expectations (Pew et al, 1977).

At some point in the modelling process, therefore, theory dictates that experimental validation should cease and design application should commence. In practice, however, model validation tends to be an ongoing, iterative process. In particular, modelling often provides the basis for a tentative system design, the adequacy of which is then checked via a hardware simulator. At the very least, the availability of such experimental data is used to adjust the values of modelling parameters if these values have not been specified a priori, i.e., 'tuning' of the model is possible. In other circumstances, more profound alterations may be carried out on the model structure itself in order to attain greater validity for the future. Possibly because the field of human performance modelling has so recently developed, instances of model validation appear to outnumber instances of model application. It may readily be appreciated that relatively few examples of modelling as an isolated systems design tool exist in the literature; rather validation and application studies tend to be found together.

3.2.2 Generality

A closely related issue to validation is model generality, as has already been implied. Due to the requirements for validation, the domain of application of models must in some ways be constrained. As a consequence, human performance models tend to vary in their applicability to different types of systems and different types of behaviour (Rouse, 1980). However, models may be ranked according to their comprehensiveness, and this is an important issue when selecting a model as a design aid (Meister, 1971b). Once again, a recurring theme is that levels of model valididty and generality should be chosen according to one's purposes. It may be unrealistic to aim towards a performance model that has high validity over a wide domain of application; however, the currently favoured solution to this problem appears to be in the use of compound or eclectic models (Sheridan and Ferrell, 1974; Rouse, 1980; Pew and Baron, 1982).

In the field of human problem-solving performance at least, there is some evidence that model valididty and generality are inversely related, i.e., a trade-off exists. That is, the most valid models are often those which have the narrowest area of application (Pylyshyn, 1978). It may be that it is possible to 'capture' human problem-solving performance, but at the expense of doing so for one subject alone, performing a single particular task, and with a certain level of exposure to the task. In the field of command-and-control performance, it is difficult to judge whether such a simple relationship between model qualities exists, although this feature has been recognised (Siegel and Wolf, 1981). Whatever the relationship, it should be noted that much empirical work revolves around extending the comprehensiveness of models in addition to refining their validity.

3.2.3 Parsimony

As discussed previously, models may contain free parameters, the values of which require estimation before performance outputs may be obtained. If theory does not specify in advance what the values of these parameters should be, then empirical data is necessary to identify the models. Such models technically have no predictive validity (and, hence, no design applicability); although they are by no means uncommon.

On the other hand, many models contain parameters that are free to vary, but which may be estimated without further empirical work. In one sense, the introduction of such free parameters may be seen as a means of increasing model generality. However, whilst such models may be predictively valid, there is a danger that they may become too complex. Parsimony is another important issue in the evaluation of models; and over parameterisation (or over-specification) subverts one of the purposes of modelling, which is the succinct explanation of performance variables (Rouse, 1980). Ideally, models should be constrained in the ratio of free parameters to variables which they predict (Pew et al, 1977). An approximate guide here is that there should be fewer free parameters than dependent variables. Hanna (1971) has also developed an index of model parsimony which is based on information theory.

The degree of model constrainment is often regarded as an index of the 'falsifiability' of the model, at least in the field of human problem-solving performance. That is, the constructs of models that lack

parsimony are difficult to test. (Hanna (1971) has added the caveat that underparameterised models may also be undesirable, due to their lack of theoretical content.) These last considerations are really applicable to the issue of the truth-value of models, that has little relevance to the subject of model usefulness during systems design.

3.2.4 Pragmatic issues

The requirements for validity, reliability, generality and parsimony should be satisfied$_2$ before a human performance model is considered as a design aid for C^2 systems. However, these qualities alone do not ensure that the model will necessarily be relevant or useful. Additional pragmatic concerns exist.

Possibly the largest concern in this respect relates to the numbers and training levels of the operators of a system. The concept of the personnel subsystem as a resource has become increasingly well-developed (Meister, 1971a), and there exists a corresponding need for models to be able to predict both the quantity and the quality of personnel who will be required in a system. In addition, the somewhat reactionary position that system design should dictate personnel requirements is rapidly becoming outmoded (Askren, 1975). Rather, personnel-related issues should be a specification of the development contract. The role of modelling should then be seen as providing a check on those specifications.

By way of anticipation, many human performance models in current use pay heed to training issues indirectly, by containing parameters that reflect individual differences in skill level. This should be seen as a minimum requirement. It is preferable that models should also make explicit the interaction of training with different types of tasks and over a period of time. On the other hand, the treatment of numbers of operators has been prominently deficient (Pew et al, 1977). Most models predict the performance of a single operator; the performance of a group of operators is then inferred by simple amalgamation. Such a procedure may neglect many important team interactions. It is theoretically possible that team structure may either facilitate or inhibit the performance of the individual, so it cannot be said that naive modelling attempts are consistently biased in their predictions.

Another pragmatic concern is the applicability of models at different stages of system development (Meister, 1971b). Design problems typically become more detailed as development proceeds; consequently models must address successively more molecular units of behaviour. Given a model which is well-suited to the prediction of performance under circumscribed situations (such as occurs during interface design, for example), it may be difficult to gather the necessary input data at conceptual stages of design. One solution to this difficulty is obviously to employ a model which treats performance in a hierarchical fashion, thus extending its applicability across stages of development.

Models with similar applications also differ in the amount of input data which they require. Generally, the amount of effort required to implement a model does not affect its theoretical acceptability, but is a significant cost in practical terms. Closely related issues are the amount of task analysis which is required before modelling, the expertise required, and the degree of computational complexity involved. These

factors are rarely discussed in a comparative manner in the literature, especially by model developers themselves.

System developers utilise models in order to assess the consequences of their designs. If the design proves to be inadequate, an alternative must be chosen by some means, often involving many subjective factors (Meister & Farr, 1966). Given a human engineering criterion, performance models rarely optimise a design with respect to this criterion (Siegel & Wolf, 1969) in the same way in which operations research models may, for example. (A possible exception is those models which deal with anthropometric data and workspace layouts.) Yet Meister (1971b) claims an important issue is the ability of the model to 'suggest' design solutions, albeit indirectly. This ability seems to depend on two factors.

First, it could broadly be said that models differ in their sensitivity to design parameters. For example, models may be constructed so that the influence of hardware configuration upon performance output is more or less explicit. An example of a design sensitive model would be one in which physical layout is a required input. In the case of modelling via a reliability data bank (that will be discussed in more detail in the next section), reliabilities may either be associated with the operation of equipment items, or with various behaviours (Meister, 1971b). The former strategy leads to more immediate design solutions, but one disadvantage is that the model then lacks generality across systems.

Second, models should also capture the nature of the human-machine interaction (Meister, 1971b). Not all models conform to this requirement; in fact, Ramsey and Atwood (1979) have delineated a spectrum of models, ranging from those that consider the operator alone to those that model the system without distinguishing the human. With regard to the former extreme, the characteristics of the system (such as operational procedures) may still be inferred as factors which both drive and limit the behaviour of the operator. However, if that behaviour is shown to be inadequate, it is preferable that the relationship between the operator and the system is explicit in order that alternatives may be designed.

This requirement suggests that it is desirable to construct models of both the operator and the system simultaneously, together with some means of interelating the two. From a human factors perspective, the model of the operator should be of greatest detail, i.e., it is sufficient to model the variables of the system which have a direct relationship with operator behaviour alone. As noted previously, such a systems model may only exist by implication, yet it is an important design feature.

The modelling of human-machine interaction is further enhanced if the models of the human and the system contain congruent terms. In practice, this means that it is most convenient to model the operator in quantitative, machine-like terms (Sheridan & Ferrell, 1974). A commonly-cited example of the difficulty which may arise from incongruent terms comes from the field of reliability engineering. Hardware designers, by convention, commonly forecast the mean-time-between-failure (MTBF) of equipment items, yet the most frequent index of human reliability is the probability of successful task completion. Whilst the latter index may have the value of allowing certain systems to be ranked according to their acceptability from a human factors viewpoint, it obscures the effects which human performance may have on total systems reliability (Regulinski, 1970).

Human-machine interaction assumes unusual importance in the design of interactive computerised systems, such as C^2 systems. In part, this importance stems from the fact that one use of the system is to extend the capabilities of the command. For example, both radar and sonar may be viewed as extensions of perceptual capability, whilst other systems may support the command's decision-making and cognitive processes. Secondly, interactive computerised systems have the characteristic of involving a dialogue, or exchange of information, between human and machine. These factors imply that the methods of systems analysis, such as modelling, should assist the designer to engineer the informational and communication aspects of the system (Nickerson, 1969). Not only must the 'physical' parameters of the system (such as keyboard layout and display legibility) be acceptable, but more abstract parameters such as software organisation need to be considered (Tainsh, 1983).

It has been claimed that models that are relevant to human performance in computerised operational systems should address monitoring and supervisory behaviour. This is possibly a minimum requirement, because such models still may not guarantee that one will be able to forecast the effects of computer software variables from a preliminary design, for instance. Whilst such a goal relating to monitoring is relatively ambitious in comparison to traditional human factors analyses, the need undoubtedly exists.

In summary, human performance models that are useful during C^2 design should satisfy both theoretical and pragmatic criteria. As with most models, they should be valid, reliable, general and relatively parsimonious; in addition, they should ideally be sensitive to operator numbers and training levels, team behaviour, system design parameters (such as hardware configuration and operational procedures), and human-computer interaction. The cost of implementing the model, in terms of personnel effort at least, may be a consideration (although from a procurement viewpoint, the onus to carry out such modelling may lie with the contractor).

4. REVIEW OF HUMAN PERFORMANCE MODELS

4.1 Types of Models

This report is primarily concerned with a particular category of model, namely, those models that can be used to assess the impact of a system design upon human performance. We are concerned with models that, in a negative sense, facilitate the identification of systems in which operator capability has been exceeded. Broadly, these situations may be defined in two ways. First, if a systems model is used, it should be possible to ascertain the relationship between human performance and system effectiveness. If the predicted human performance is unable to maintain system effectiveness to a specified level, then it may be said that the operator has been identified as a weak link in the system and that re-design is necessary. Alternatively, it may be possible to forecast the required operator performance alone, and then to determine whether that expected level of performance is unreasonable on a priori grounds.

By way of anticipation of what follows, the latter approach appears to be more widespread, i.e., fewer models incorporate the effects of operator performance upon system effectiveness. (This observation is particularly true of the 'bottom-up' approach to modelling.) Thus, human factors specialists are frequently concerned with models which predict the levels of operator performance that are required by the demands of the system. It then remains to decide whether that performance is unreasonable, i.e., whether operator capability has been exceeded or, in more common terms, whether workload is too high.

Generally, workload is a concept that is open to many interpretations. From a modeller's viewpoint, workload is most frequently defined as the percentage of time for which an operator is occupied on a particular task. Models which forecast time-on-task are therefore particularly relevant. If the time constraints for a particular task may be estimated, and operator performance time is predicted to be larger (or comparable) to that estimate, then system re-design is suggested. An alternative class of models that are also relevant to the concept of workload are those which forecast human reliability. In a frequentistic sense, if a task is forecast to be performed unsuccessfully for a relatively large percentage of attempts, then workload may be said to be excessive by implication.

The most popular concepts of workload in modelling are therefore based on two parameters of human performance: speed and accuracy. Unfortunately, there is some doubt whether these parameters adequately characterise performance on non-manual, cognitive tasks. For example, within C^2, it is often the quality of decisions which is regarded as the critical index of performance, and for which design innovations are sought. Decision quality may be related to speed and accuracy of performance, but only in an indirect sense. It is preferable that models should incorporate such abstract indices of performance in order that the cognitive demands, or cognitive 'workload' of systems may be predicted. Anticipating once again observations made later in this chapter, it can be said that this goal is yet to be realised.

In this report, we have excluded a large number of models due to the fact that they fail to satisfy the evaluative criteria that were

discussed in Section 3.2. One popular class of models, namely, human problem-solving models, may be excluded on a number of grounds. Much work in this area has revolved around construct validation of the model against the verbal protocols of a subject. The focus is not on making performance forecasts, least of all in a quantitative fashion. The models in this class also tend to be extremely task-specific and thus tend to lack generality.

A second class of models that have been excluded are those which are concerned with personnel allocation. That is, given a personnel resource with certain characteristics, these models may be used to assign personnel to tasks in a fashion that optimizes manpower usage. These models are rarely concerned with performance forecasting per se, and thus have not been considered. However, one model in this report, namely that of Siegel & Wolf does address this issue indirectly.

Finally, we have generally excluded models which predict the maintainability of systems. Whilst some of these models may forecast human performance in a quantitative manner, the review is restricted to systems operability for reasons of convenience. Reviews of maintainability models may be found in Smith, Westland & Crawford (1970) and Meister (1971b).

For the models that have been included in this review, Pew & Baron (1982) have made a useful distinction between those that are psychologically-based and those that have arisen within an engineering domain. As will be seen, the methodologies of these two styles of modelling are profoundly different. The so-called psychological models are characterized by the fact that they treat individual tasks as the basic unit of analysis. They include the relative profusion of network models and a smaller number of information processing models. The engineering-based models are characterized by their treatment of the human as a component of the system (described mathematically). They include models derived from estimation, control and queuing theory.

4.2 Network Models

A number of models may be subsumed under the category of network-based techniques. All have a number of features in common. First, network modelling requires that system performance is decomposed into a number of tasks at a convenient level of analysis. System diagramming often aids the visualisation of task relationships from a conceptual design. Human performance data for each task, such as average completion time and average probability of successful completion, must be derived by some means. Total performance is then predicted by aggregating the individual task data according to a set of rules or procedures.

A fundamental distinction within the various means of aggregation has been claimed to be that between analytic and simulation methods (Meister, 1971b). The analytic approach relies on the use of combinatorial statistics; a frequent convention being that the completion times of independent tasks should be added whilst their reliabilities should be multiplied. (The use of the term 'analytic' in this context is somewhat misleading, because efforts to aggregate task data into an overall prediction actually constitute a synthetic operation.) The simulation approach requires that the anticipated task sequence is repeatedly exercised (usually in fast time) in order to obtain the

performance outputs. As the individual task parameters are stochastic in nature, a Monte Carlo procedure of selecting from the parameter distributions is often used. Different simulation runs may therefore be quite disparate in their predictions; however, task performances eventually should converge to their expected values with repeated exercising of the model.

The simulation approach mirrors reality better than the analytic approach, for at least two reasons. First, if each simulation run represents the performance of a single operator, performance on every task will not be identical to the average performance. That is, the modelling of human randomness (both within and between operators) is accommodated. Secondly, prediction via the so-called analytic method presumes a fixed number of tasks, which narrows the domain of application of the model. Simulation methods may accommodate sequential task variability through the use of precedence relationships between tasks, the details of which will become clearer as the particular models are discussed more fully.

4.2.1 Human performance data banks and reliability trees

As mentioned previously, all network modelling techniques require individual task performance data as an input. The scarcity of these data has often evoked concern (Meister, 1967), particularly with respect to human reliability, i.e, the probability of successful task completion. This has led to efforts to compile performance data banks for future use. As it transpires, however, the organisation of these data banks tends to imply a human performance model, making it even more appropriate that data banks should be discussed within the topic of network models.

Aside from their intrinsic value as inputs for network modelling, data banks may be used to predict system performance by the so-called analytic method (Meister, 1971b). That is, the individual task performance data may be aggregated by some form of combinatorial equation. Most desirably, the initial systems analysis is carried out at such a level that the individual tasks may be considered to be independent. In other words, performance on any one task should not depend upon the particular sequence of tasks in which it is embedded. The practical consequence of this requirement is that systems analysis should be carried out at a relatively molar level. It is then convention to add successive task times whilst multiplying task reliabilities in order to estimate the values of these two parameters of a systems level. The reliability of parallel tasks is taken to be the minimum overall.

Two types of organisation of data bases exist which reflect different philosophies of systems analysis (Payne & Altman, 1962). The first considers a range of behaviours that are closely linked to equipment type. For example, performance times and reliabilities for actuating switches, reading dials, etc. have been collected together. The second type of organisation uses the task itself as the unit of analysis. Basic categories may be 'inspection' 'assembly', etc. and appear to be relatively molar. It may therefore be easier to avoid a requirement for representing task interactions when using this style of data organisation. Due to its psychological orientation, the data are also of more use as an input to network models which address cognitive behaviour. On the other hand, there is a problem that hardware engineers may require molecular performance data in order to assess their designs (Pew et al, 1977). As a consequence, design solutions may be more

immediately apparent when using equipment-related performance data
(Meister, 1971b), provided that these data have been interpreted
legitimately.

The standard definition of reliability as 'average probability of
successful task completion' means little by itself. For one thing, the
statistic is a point-estimate; it fails to account for human
variability. The method of sampling from a reliability distribution, as
is done during simulation modelling, is preferable.

Secondly, there is a distinction between the use of error data
for evaluative and predictive purposes (Pickrel & McDonald, 1964). That
is, alternative systems may be ranked according to their composite human
reliability indices, as part of an evaluative program. However, if one
wishes to predict the actual frequency of occurrence of some task error,
more information is needed regarding the conditions under which the task
will be performed. For example, it may be that an operator makes an error
and then corrects it, or correction may occur through the agency of a co-
operator. The performance of multiple tasks may result in error
interactions. Successful task completion may also be a function of the
time constraints imposed by any system. What is an error in one system
may be regarded as mere slow performance in another.

Thirdly, human error data on their own give no information about
the significance of that error. The failure of the total system is of
paramount importance, so some means of assessing the effects of 'human'
error on total system performance is needed.

Lastly, task interactions are an ever-present problem when using
analytic reliability models. Despite efforts to analyse the system into
independent tasks, the performance of these tasks as a whole may be
different from the aggregation of the individual performances. Task
interactions may be both facilitatory and inhibitory in nature; in either
event, the validity of the reliability model suffers.

Some attempts to overcome these conceptual problems of
reliability will be illustrated in the following discussion of particular
models.

(a) Technique for Human Error Rate Prediction (THERP). THERP (Swain,
    1964) uses a tree-structure approach to reliability prediction.
    First, systems analysis is used to identify tasks that are
    critical to the mission. The relationship between these tasks is
    then represented in a tree format. That is, each task forms a
    node in the tree. At each node, there may be a number of
    possible outcomes. Commonly, a dual branch exists, representing
    the outcome of either success or failure of that task. Tracing
    along one branch of the tree represents a particular event
    sequence. Human reliability data are posted to the nodes of the
    tree and are aggregated in the standard fashion. The technique
    allows an estimation of the overall reliability of either any one
    task sequence or of all possible sequences. In addition to the
    prediction methodology, concurrent efforts have been made to
    establish and refine a data bank.

    Some relatively sophisticated features of THERP distinguish it
    from other analytic reliability models. The first is that non-

independent task data may be aggregated through the assignment of conditional probabilities to the nodes of the reliability-tree. Naturally, this method presumes some expertise on the part of the analyst. There is also a facility for adjusting the reliability of operators in a group situation.

The model appears to be somewhat more systems-oriented than other related approaches. The probability that unsuccessful task completion will result in system failure may be specifically incorporated into an aggregation rule. Hardware failures may also be aggregated alongside human reliabilities, i.e. the model represents human-machine interactions. (This representation presumes that hardware reliability data will be available as a function of trials, rather than as a function of time, which is more common).

The refinements of THERP make it the most acceptable of the analytical reliability models on both theoretical and pragmatic grounds. However, a comparatively large onus is then placed on the modeller to gather input data. Whilst the compilation of large data banks may reduce this effort, others (Knowles, Burger, Mitchell, Hanifan & Wulfeck, 1969) have claimed that a more practical method is to use expert ratings anew for each system.

(b) <u>Pickrel and McDonald model</u>. The model of Pickrel and McDonald (1964) was not actually presented in a working state, but was one of the early examples of how human performance could be quantified during systems design. It is presumed that systems analysis yields an allowable time for completion of the individual tasks, based on mission requirements. Total performance time is predicted by simple addition of task times. If estimated task times are higher than those allowed, operators are said to be under excess workload.

As for reliability, task error is defined as failure to complete the task within a given time period. The probability that the system would be degraded by the occurrence of that error is determined on a judgemental basis, as is the severity of any degradation.

The model has not been validated, so no assessment may be made of its utility. The model assumptions are closely aligned with those of other (validated) reliability models, so its use as a simple and approximate predictive tool appears reasonable.

(c) <u>Sandia Human Error Rate Bank (SHERB)</u>. Another probability-tree method by the developers of THERP is SHERB. The distinction of the latter approach is that data are organised according to psychological dimensions, rather than according to the operation of items of equipment (Pew et al, 1977). The data are useful as an input for modelling of more cognitive processes. The molar level of organisation may also reduce the calculations necessary to accommodate task interactions.

One may discuss both the validity of data banks themselves and the validity of analytic reliability models. In the case of data per se, much of it has been derived from laboratory studies and then extrapolated

to human performance in systems. Direct extrapolation is only valid if the real-world tasks match the laboratory tasks exactly. Hence, modification of reliabilities by expert judgement often occurs. The validity of the network model itself is a further issue.

A frequent convention of simulation network modelling is to assume a normal distribution of task times and reliabilities, based on the Central Limit Theorem. One study (Mills & Hatfield, 1974) has shown that this approach may be mistaken. That is, small errors of prediction are likely to occur if the actual distributions are not used. The assumption of normality has also been challenged by Bradley (1975, 1983).

The same study by Mills and Hatfield affirmed the logic of summating independent task times in order to arrrive at an overall prediction. However, the method of multiplying reliabilities, even with tasks which seemed to be independent, was shown to be erroneous for the range of tasks considered (which basically involved human computation). This last result undermines the validity of many analytic reliability models, although those which utilize conditional probabilities, such as THERP, may be exempt. Unfortunately, it is difficult to assess the validity of the various reliability-tree methods as few test studies exist.

As discussed, the major contributions of data banks to design are as an input to further network models. Few design applications of reliability-trees per se exist in the open literature.

4.2.2 Siegel and Wolf naval models

The model of Siegel and Wolf (1961) represents the first attempt to model psychological processes via a simulation network. As with all network models, an initial systems analysis is required to define the tasks which constitute the network. The basic task parameters are average performance time, and its standard deviation, and average probability of successful completion. These values may be derived from a performance bank, if available, otherwise they must be estimated. Task performance time is presumed to be normally distributed, which may not be a valid assumption under all circumstances. On any simulation run, the actual performance time is pseudo-randomly selected from the distribution of performance times which is defined by the mean and standard deviation. Over a large number of simulations, it is expected that mean simulated performance time for any task will converge to the input value. Total performance time for any simulation is computed by the summation of task times in the familiar manner.

Probability of successful task completion is also subject to a pseudo-random selection process. During the simulation exercise, however, task reliabilities per se are disregarded in favour of a binary decision, namely, whether the task may be considered to have been completed successfully or not. If so, simulation of the next task occurs. If not, simulation of the original task may be repeated,, depending upon other factors. For any simulation run, the model does not output an aggregate probability of success. Rather, total performance time is compared with some standard in order to decide whether the mission was successful or not.

In contrast to the models implied by reliability-trees, the Siegel and Wolf method of reliability prediction avoids combinatorial statistics. According to Meister (1971b), this binary nature of reliability assessment is superior because the problem of mathematically accounting for task interactions is also avoided. On the other hand, if one wishes to assess the composite reliability of tasks that do interact, the model offers no solution to this problem.

Precedence relationships between tasks also exist. The inputs to the model require an indication of which tasks are essential to the completion of the mission. The significance of this information is that essential tasks must be repeated in the event of failure. Non-essential tasks may be bypassed if time constraints apply. To achieve these strategies, an indication is also needed of the next tasks to be performed in the event of either success or failure of the original task.

As mentioned, the allowable performance time of the mission is an important variable. It is presumed that the operator has the ability to decide continuously whether the remaining tasks will be performed on time, given his/her average performance rate and an assumption of no further repetitions. If it is calculated that insufficient time remains for completion of essential tasks, then stress conditions are simulated. Stress is defined as the ratio of time required to time remaining.

Siegel and Wolf's introduction of a stress concept was a valuable psychological addition to network modelling. Up to a certain point, stress is regarded as facilitating performance. That is, the value of mean performance time decreases, whilst probability of successful task completion increases. At a certain point, which is defined as the threshold, stress begins to have inhibitory effects, analogously to the familiar inverted U curve of performance. This degradation eventually reaches a constant value if stress becomes great enough.

The value of the stress threshold represents an operator characteristic, thus allowing the modelling of individual differences, which is another relatively sophisticated psychological concept. Generally, stressful conditions are simulated by suitable reductions of allowable mission time. This may have either facilitatory or inhibitory effects overall. If the latter, the effects may be counter-balanced by modelling operators with larger stress tolerances. In addition, a second individual difference parameter is operator speed, a value of which is required in the input. Lower values cause expected task performance time to decrease, which also increases the probability of the mission being completed on time. One limitation to the use of this parameter is that it is presumed that faster operators are faster on all tasks. If the tasks in the network require differential abilities, the latter assumption may not be valid.

The original model accommodated single operator performance only for each run. A later modification (Siegel & Wolf, 1969) addressed both dual-operator and group performance. Both models have formed the basis for a number of studies in the last 15 years. Two models in particular have been well validated. These are the 1-3 man and 4-20 man models which have been designed for reliability prediction within the U.S. Navy. The 1-3 man model is similar to the original single operator model and will not be described further, in favour of a description of the 4-20 man model.

Briefly, three classes of variables are required as inputs, (along with the standard data such as allowable mission-time, etc, and also a factor of sea-state). The first class defines the characteristics of the personnel. For example, the modeller must specify the number of men who hold each rank, and must give an indication of their training specialty. Work pace, stress tolerance, fatigue and aspiration level are all parameters which modify task performance. In addition, the physical capacity of the personnel must be defined through assignment of values to a number of variables. A more complete list of these variables may be found in Table 5.

The second category of input data relates to the equipment which will be used on the mission. Most crucially, the failure rate and average repair time must be specified for each event.

Number of men holding each rank, and a training speciality code

Body weight of each crew member, and the standard deviation of the crew

Average proficiency level in primary and secondary specialty

Average work pace

Average stress tolerance threshold

Average caloric intake

Average duration of incapacity

Average number of hours since sleeping

Minimum fatigue necessary for sleep

Average physical capability

Average capability after a full work-day

Average short-term power output

TABLE 5   -   **Input data for the Siegel & Wolf 4-20 man model (personnel variables for each crew member)**

Lastly, the events which comprise the task network must be described parametrically. In a similar fashion to the original model, mean task duration and essentiality must be defined. In addition, an indication must be made of both the number and type of personnel who are required to perform each task, as well as the energy demands involved, and the 'mental load'. Table 6 has a more complete list of the equipment-related and task-related input variables.

Equipment Variables

Failure rate of each item

Average repair time, and the standard deviation

Number and type of personnel required to repair each item

Mental load of repair

Consumable use

Event Variables

Mean duration of each task, and its standard deviation

Relative essentiality of each task

Mental Load

For each consumable, the rate of expenditure and the
minimum amount necessary to perform the task

Energy demands

Hazards encountered

Number and type of personnel required

**TABLE 6 — Input data for the Siegel & Wolf 4-20 man model
(equipment and event variables)**

The simulation then proceeds through a number of routines. The
first stage is crew formation. That is, although average personnel
characteristics have been specified in the input, the program generates
further mission-specific variables relating to aspiration, competency,
etc. Equipment failures are then determined. Under the presumption that
failures are randomly distributed, the time sequence of their occurrence
is generated. It may be seen that a distinction is made between the
scheduled events which have been defined in the input data, and other
'unscheduled' events. Personnel are then assigned to both types of events
in a manner that automatically selects the most highly qualified and able
people who are available. Unessential events may be by-passed if either
time is lacking or if personnel are unavailable. Unscheduled events may
not be by-passed.

Monte Carlo simulation is used to obtain event completion times
in the standard fashion (subject, of course, to the modifying influence of
many of the input parameters just described). Total simulated mission
time may then be obtained; however, the group mode: is largely concerned
with less gross dependent variables. In particular, an effort has been
made to have the model's outputs conform to the terms used within
traditional reliability engineering. A listing of these terms, together
with the associated definitions, is given in Table 7.

Briefly, forecasts are made of both hardware and human
reliability, along with a combined systems reliability index. Hardware

reliability is defined by four commonly-used variables: reliability, availability, mean-time-between-failure and mean-time-to-repair. The performance of the crew is then stated in what are believed to be congruent terms: reliability, availability and mean-time-to-repair. The latter two variables are obtained by giving human performance a somewhat unusual interpretation, viz. human availability is related to the number of occasions on which 'events' require attention but are necessarily ignored (e.g. through excess workload), whilst human mean-time-to-repair is related to the amount of time used to repeat tasks after an initial failure.

Human reliability $\quad (1 - \dfrac{\text{number of task failures}}{\text{number of attempts}})$

Human availability $\quad (1 - \dfrac{\text{time lost in task repetition}}{\text{total mission time}})$

Human mean-time-to-repair $(1 - \dfrac{\text{time for task repetitions}}{\text{total number of repetitions}})$

Equipment reliability $\quad (1 - \dfrac{\text{number of failures of each item}}{\text{number of simulated runs}})$

Equipment availability $\quad (\dfrac{\text{up time}}{\text{up time and down time}})$

Equipment mean-time-between-failures

Equipment mean-time-to-repair

System reliability $\quad \dfrac{\text{number of equipment failures + number of task repetitions}}{\text{number of simulated runs}}$

System availability (Human availability x Equipment availability)

System mean-time-to-repair

TABLE 7 - Outputs of the Siegel & Wolf 4-20 man model

   In addition to these standard outputs, a facility exists for performing a finer analysis. For example, the types of tasks which have been failed most frequently within the simulation may be tabulated. Alternatively, individual task performance time may be considered to be of greater interest and may be tabulated. The average stress per task may be calculated. Task failure may also be calculated as a function of particular members of the hypothetical crew.

   Siegel, Leahy & Wolf (1978) have indicated explicity that 'trade-off' runs are an important aspect of their modelling. That is, values of the input parameters may be systematically varied in order to observe and

compare system performance of interest. For example, operator speed or aspiration level may be manipulated in order to note the contribution that changes in these variables make to mission performance. From a pragmatic viewpoint, the most significant trade-offs are possibly those involving manpower or training aspects. Generally, the model structure presumes that personnel numbers and training are factors which act as limits upon mission performance. If a small increase in those factors is forecast to yield relatively large increase in system effectiveness variables, for instance, then the simulated trade-off may be used to justify some personnel planning policy for the system under consideration. Alternatively, given system effectiveness criteria, the modelling may be used to make projections of personnel requirements.

In principle at least, trade-offs may also be applied to hardware reliability. As equipment repair is presumed to deplete personnel resources that would otherwise be available for scheduled tasks, it may be possible to assess the value (in a systems sense) of 'upgrading' the equipment. It should be noted, however, that this use of the model, whilst addressing maintenance aspects of the system, does not allow one to make trade-offs regarding hardware operability. In that case, it would be necessary to perform systems analysis and modelling for each hardware configuration in order to make comparisons between designs.

4.2.2.1 Validation studies

The validation studies of the Siegel and Wolf models are probably the most accessible of any that are relevant to C^2 performance. Generally, validation of the group model has been less rigorous than that of the single operator model, due to the fact that they yield a 'coarser' analysis of the human-machine system (Siegel, Leahy & Wiesen, 1977).

The predictive validity of the single operator model has been tested against empirical performance in two situations: landing onboard an aircraft carrier and a missile launching task. Records showed that 50 out of 81 landing attempts had been unsuccessful in the past. Systems analysis revealed a basic sequence of 37 tasks, as well as a maximum allowable performance time. The model was then exercised 81 times using particular values of the input parameters. It would found that the external failure rate could be predicted using operators of average speed and a certain stress threshold. Similarly, the performance of the missile launching task could be predicted, given operators of the same stress threshold but above-average speed. That is, the average stress threshold of the two groups of operators was presumed to be the same, although no justification was given.

Two criticisms of the theoretical acceptability of the Siegel and Wolf (1961) model emerge from these validation studies. The first is that the model technically had no predictive validity in these circumstances, because adjustment of free parameters was required to fit the model to the data. In other words, the model lacks parsimony, although it is presumed that more empirical work would allow specification of the values of these parameters in advance of testing. The fact that operator speed was regarded as a fixed parameter during one test and not during another is also somewhat inconsistent.

Secondly, no details were given of the systems analyses that led to the contruction of the model in these two studies. Presumably,

operational systems were available on which to conduct analyses. If so, the validity of the model has not been established under conditions in which the task network has been derived conceptually. It is possible that the model was validated against the same data set that it is supposed to predict, which limits both validity and its use as a design tool.

The group model (Siegel & Wolf, 1969) was tested against data from a 21-day submarine training mission. Performance-related variables and predictions of crew composition were found to correlate well with empirical data, although adjustment of free parameters was once again necessary. Social variables were also claimed to be predicted well; however, the testing procedure might have lacked rigour due to the fact that the model's predictions of crew morale, cohesiveness, etc were given to observers to be rated, rather than the predictions being compared with empirical data.

The 4-20 man model has been studied for its ability to simulate human performance within the AN SQS-26 sonar system. Siegel,Wolf & Williams (1976) gathered the validating data by interviewing experienced officers within the system about a number of aspects of crew performance. The topic of the interview was not a particular mission in which they had participated; rather, a representative scenario was devised with the help of the appropriate personnel. The officers were then required to give estimates of five factors, viz:

(i)   percentage of tasks successfully completed on first attempt,

(ii) percentage of time spent on normal duties,

(iii) percentage of time spent on repair duties,

(iv) percentage of tasks necessarily ignored,

(v)   average degree of fatigue experienced.

The scenario in fact consisted of seven variations. Basically, various combinations of crew proficiency and manning level were hypothesised, so that differential effects on the five performance variables could be judged. Presumably, the actual task sequences were the same in each scenario in order to facilitate a comparative analysis (although this aspect is unclear from the report).

The scenario then formed the basis for construction of the model network. The values for a number of input parameters, such as task completion times, were once again derived with the aid of the operational staff. The proficiency and manning variations were achieved by adjustment of the appropriate model parameters.

As there were five dependent variables and seven conditions, a total of 35 indices were available for validation purposes. The correlation between the model's outputs and the estimates of the sonar personnel were then calculated. Generally, these correlations were statistically significant, although the forecasts of repair time and fatigue experienced were regarded as unsatisfactory for a number of reasons. Overall, the model was judged to be valid.

On the other hand, some criticism may be made. The first point is that it is questionable whether a demonstration that the model's outputs correlated with the criterion data above chance-level is

sufficient evidence for validity. The second concern relates t: the use of subjective estimates. As discussed elsewhere in the report, such data may be reliable and virtually essential in many circumstances. However, no indication was given in the study of Siegel et al (1976) about the reliability of the estimates, nor about the formality of the procedures used for collecting those data. On the positive side, albeit indirectly, the model has been regarded as sufficiently well-validated for design applications to be made in later work.

4.2.2.2 Design applications

Both the 1-3 man and 4-20 man models have been used to make recommendations for the design of Naval sonar systems. The 1-3 man model has been applied to the AN/SQS-26 and AN/SQR-10 system, whilst the 4-20 man model has been applied to the AN/SQS-26, LAMPS and AN/SWR-19 systems.

The 1-3 man model (Siegel, Leahy & Lamb, 1976a) was exercised through the simulation of a 24 minute scenario. Briefly, two operators were involved. The mission goals included detecting a target vessel and hence performing a change of course. Different conditions were simulated in which both operator speed and stress-tolerance varied. The model outputs that were analysed included failure rates (for each task), task repetition times, average degree of stress and average amount of time that task initiation was delayed.

Two major findings emerged from the study. The first was that a hypothetical 'target reacquisition' task was performed most unsatisfactorily of all tasks and degraded system effectiveness. It was suggested that an augmented (predictor) display might alleviate this problem, although it is unclear how this conclusion was made. Secondly, only those operators of superior proficiency were forecast to perform at a satisfactory level. Thus, it was suggested that a training programme might be necessary in order to increase the average proficiency level.

The 4-20 man model (Siegel, Leahy & Lamb, 1976b) simulated a hypothetical four-hour mission in which an enemy target was attacked. The sonar system requires four operators and some degree of team co-operation. For example, data from two operators must be 'merged' before a change of course may be selected. This interaction is modelled via precedence relationships. That is, it may be essential for one operator to complete a certain task before the team may continue with its activities. Unsuccessful task performance at an individual level thus increases the 'waiting time' of the team and generates time-stress. Other non-specialised tasks may be performed by whichever member of the team is available, through the personnel assignment routine.

Variations on the basic scenario included conditions in which pace, aspiration, proficiency and leader expectation were manipulated. The latter factor is generated by the program and once more represents a team-related variable. Discrepancies between the leader's expectation and the team's average level of aspiration modify task performance.

The major conclusion from this study was rather gross, namely, that system effectiveness was limited more by crew performance than by equipment reliability. Further, increases of crew pace or aspiration were forecast to have a relatively small influence. In practical terms, this result suggested that the institution of a training programme would be an

ineffective means of increasing performance, and that fundamental system re-design was necessary.

In particular,the hypothetical supervisor of the sonar operators was identified as an unreliable link in the system. Decision-aiding of an unspecified nature was seen as one solution to the problem. Other design recommendations included improving both the communications network and some of the human-machine interfaces. A redistribution of tasks amongst the team was also suggested.

Generally, it may be appreciated that both Siegel and Wolf models have relatively great power for suggesting design solutions. This is largely because the detail of the simulation models' outputs may allow deficiences within the system to be identified. However, the precise means for rectifying those deficiencies may still not be apparent. For example, decision-aiding may be a reasonable solution in some circumstances of excess operator workload, but the form which that aid should take may remain obscure. This is because background research has merely indicated that a particular cognitive task is subject to error, which is then translated into the input data. Similarly, if an increase in operator ability is suggested, it may be difficult to relate the values of one or two skill parameters to the actual level and type of training required (or to more rigid selection criteria). In contrast to analytic reliability models, however, simulation models are generally superior for facilitating systems design. The design applications of the Siegel and Wolf models in particular have probably been better documented than any other network model.

The 4-20 man model is also distinctive in its treatment of group variables. The trade-off studies between crew numbers and system effectiveness address a very pragmatic need within modelling. Team organisation has also been addressed to a limited extent, which is desirable. That is, the personnel assignment routine of the model allocates crew to certain tasks from a pool of available personnel. If the characteristics of that pool are varied (for example, by changing the ratio of officers to lower ranks), then the effects on system performance may be forecast. On the negative side, the results of such a manipulation are trivial in a certain sense, because the model presumes that the greater the average mission-specific proficiency of the personnel pool, the more optimal is system performance. However, it would still be possible to forecast the minimum personnel characteristics that would be required for a certain level of system performance.

The Siegel and Wolf Naval models simulate the hardware component of the system to the extent that unscheduled equipment maintenance duties may increase the time-stress of the primary mission. As for system operability, that factor is reflected in the performance data for each task that is a required input of the model. By this method, it is unlikely that human-computer interaction could be modelled in great detail.

The effort required to implement the Siegel and Wolf model is relatively high in comparison to analytic prediction methods. One reason for this is that precedence relationships must be visualised in the original task analysis. However, Siegel, Leahy and Wiesen (1977) noted that the time required for devising the task network (from a scenario) is generally less than that required for devising values of the input parameters. Lastly, any simulation method is likely to require relatively large amounts of computer programming and operating time.

How applicable the Siegel and Wolf model is in the general conceptual stages of design is difficult to assess. On the positive side, the model makes no presumptions about the level of detail of the tasks which constitute their networks. If a task sequence may be defined (however molar), and task completion times and reliabilities may be assigned, then the model may be implemented. However, Meister (1971b) at least considers that suitable input data may be difficult to find at early stages of design. This comment would apply particularly to the more complex group model. It is worth noting that both design applications of the model which have been discussed in this report occurred part-way through the detailed design stage.

4.2.3 NETMAN

NETMAN has been developed specifically to aid the on-going design of a field exercise management system for the U.S. Army. It is somewhat distinctive in that the modelling has been oriented towards an information system/communication set in particular. Many of the model's concepts are a heritage of the work of Baker (1970), in which information flow within the Tactical Operations System (TOS) was analysed. NETMAN itself has been developed by Applied Psychological Services, and has some relation to the Siegel and Wolf (1961) model; ie, it is a digital network simulation which places significant emphasis on the effects of time-stress upon human performance.

In contrast to the Siegel and Wolf models, NETMAN is more of a systems model. That is, system effectiveness is seen as a combination of both personnel characteristics (such as numbers and skill level) and characteristics of the information which is being managed (such as number and length of messages).

Briefly, the structure of NETMAN presumes that a number of messages are generated in the field and then compete for processing. These messages must pass through three hierarchical levels within the system that are staffed by different personnel. Messages are first received by one of nine referees, who then transmit the processed data to one of nine radio operators, whereupon they may be processed by a computer and directed towards a central controller. At each level, a number of standard tasks exist. Other communication loops also exist: the controller may initiate new messages for the referees, whilst the referees are also in contact with each other.

Inefficient message processing at all three levels of the system leads to the formation of information 'bottlenecks'. That is, unattended messages form a queue to await processing. Some messages may be assigned greater priority than others by parametric description. As the components of the system are interrelated, ineffective performance at one point may modify the conditions of performance at another. The model is thus sensitive to the organisation of the system. For example, the ratio of operators at each hierarchical level may be manipulated in order to forecast subsequent system effectiveness.

The behavioural input parameters of the model include operator speed, precision and aspiration. The informational input parameters include the number of field messages, the amount of work which the later messages generate, and the length of these messages. Needless to say, the

model structure presumes that the greater the 'traffic' within the system, the more the operators will be placed under time-stress. This factor then modifies both task performance time and human reliability, the average values of which have been established by a previous systems analysis.

The outputs of the model are primarily systemic in nature. In particular, four variables are considered to be indicative of system effectiveness. These are 'thoroughness' (number of messages completed: number of messages received), 'completeness' (average number of successful tasks), 'responsiveness' (message processing time: message handling time) and 'accuracy' (amount of information lost). These indices are calculated for an appropriate length of simulated time. As with other Siegel and Wolf models, various conditions may be created by altering the personnel characteristics of the model.

4.2.3.1 Validation studies

The validity of NETMAN has been considered to be sufficient for an application study to be made (Siegel, Madden and Wolf, 1981). However, the details of the validation attempts have not been obtained at this stage. A sensitivity analysis was performed (Siegel,Leahy and Wolf, 1977) in which the internal consistency of the model was established.

On the negative side, a logical analysis of NETMAN suggests that its generality is limited. For the model to be validated against operational data, there would be a requirement for an information system with rather unique characteristics, namely,a three-stage hierarchical communication net. This critisism is not profound when one takes the purpose of NETMAN into account, ie, to model a particular field exercise management system. However, in order to model other information systems, a totally new model would be required.

4.2.3.2 Design applications

Whilst NETMAN was originally developed to model an operational information system (TWSEAS), it has more recently been used to assess the desirability of modifications to that system. In particular, the Exercise Monitoring and Report System (EMARS), which is in the conceptual design phase, has been proposed as an addition. The details of the difference between the two systems are unclear from the study concerned (Siegel et al, 1981), but it appears that EMARS is more highly automated. A simulated comparison of the two systems did, indeed, suggest that EMARS has superior effectiveness, particularly under conditions of frequent or lengthy message-handling.

Generally, the pragmatic qualities of NETMAN are good. The model is sensitive to operator numbers, skill, system organisation and procedures. It is specifically a computerised system model, which is rare but certainly timely. On the other hand, the structure of NETMAN would probably have to undergo considerable modification before it could be applied to C^2 design.

4.2.4 Systems Analysis of an Integrated Network of Tasks (SAINT)

SAINT may be regarded as a logical development of the Siegel and Wolf (1961) model. It is a network simulation model once again, and so all the comments that have been made regarding that general methodology

still apply. However, SAINT is distinguished by having greater flexibility in the model structure which allows the introduction of more varied psychological contructs, and increases the domain of application.

Once again, task performance is defined by the two variables of duration and reliability (Pritsker, Wortman, Seum, Chubb and Seifert, 1974). In contrast to the Siegel and Wolf (1961) model, however, it is not routinely presumed that task times have a normal distribution; in fact, a choice of 11 standard distributions exist whilst there is also a facility for specifying alternative distributions through subroutines. Two parameters exist that reflect proficiency level: operator speed and accuracy. The values of these parameters influence the pseudo-random selection procedure from the distribution of the respective performance variables. A 'moderator function' allows the value of such individual difference parameters to be varied at any time within the task sequence of any one operator. This facility partially solves the problem of modelling task interactions; for example, fatigue effects may be built in.

The concept of operator stress and stress threshold is used in an almost identical manner to that of the Siegel and Wolf (1961) model. Stress may not necessarily be task-related; for example, Pritsker et al (1974) give preliminary details of the concept of environmental stress, such as results from ionising radiation. A second major performance variable is goal gradient, that provides for a 'commonly observed' increase in performance accuracy as the completion of a group of tasks becomes closer.

Single, dual or group task peformance may be modelled. In the case of more than one operator, a cohesiveness parameter is available that modifies the aggregate performance. The task network may therefore be simulated using different numbers of operators in order to assess the effects on overall performance. The output of the model allows performance to be classified by task type, which is convenient for purposes of analysis.

The precedence relationship between tasks within SAINT are more versatile than those in the Siegel and Wolf (1961) model. The next task may be selected on a purely random basis, in order to model situations where there is no fixed task sequence. Alternatively, the selection of a task may be conditonal upon the completion of other tasks or the satisfaction of other performance conditions.

SAINT combines human performance and system behaviours into the one structure. In particular, the expected states of the system are required as inputs for the model. The values of these states determine the moment-to-moment activities of the operators to some extent, i.e. human peformance is regarded as dynamic. For example, flight path equations, etc, must be derived as a function of time and input when modelling the control of remotely piloted vehicles (Miller et al, 1978). Analytical human performance sub-models may also be incorporated.

The flexibility of SAINT means that is it technically less of a model per se and more of a framework around which proposed models may be described. Through the use of subroutines and modularisation, minimal constraints are placed on the modeller beyond the necessity of analysing the system into a task network. Whilst Siegel and Wolf's models demand input values for specific parameters, the structure of SAINT is not

fixed. On the other hand, SAINT provides no guidance to the user about the construction of new performance models. The standard parameters which have been described may therefore be very welcome in some circumstances.

4.2.4.1 Validation studies

SAINT has commonly been validated against data from human performance in simulators in preference to data from operational systems. In most cases, this method appears to have been favoured due to both the lack of an operational system and also the convenience of the controlled conditions which may be created in a simulator. Data from simulators has also frequently been used to estimate the parameter values of the model.

SAINT has been applied to evaluation of the U.S. Air Force's Digital Avionics Display System (DAIS) (Kuperman, Hann & Berisford, 1977). In a simulator, pilots were required to perform a primary task of flight control whilst performing a secondary task of 'multi-function keyboard switching' that is a component of the DAIS. The main dependent variable of the model was time to achieve control of the plane from some set of disturbed initial conditions, and it was claimed to be predicted satisfactorily. In addition, the model showed that the level of difficulty of each task affected performance and that the two tasks did not interact with each other, which was also verified in the simulator. The validation was weakened somewhat by the fact that eight parameters relating to the pilots' sensitivity to flight control information, plus their weightings, could not be estimated in advance of empirical data.

Wortman, Seifert and Dukert (1976) used SAINT to simulate both operator and system performance in a remotely-piloted vehicle control task. Although six operators were involved, the task was largely one of individual performance by each operator. Once again, a prototype of the system was available before the modelling exercise. The objective was to 'duplicate' the real-time performance, in order to demonstrate the applicability of the model. Operator performance variables related to number of control actions, such as velocity and heading changes; whilst system performance was measured by deviations from a pre-defined flight path. It was found that 265 out of a total of 281 of the simulated dependent variables were within an acceptable range.

The investigators made it explicit that data from the real-time simulation were used to alter both the structure and the parameter values of the SAINT model, until sufficient agreement was found. On the other hand, SAINT was used to locate 'inaccuracies' in the real-time simulation by suggesting which independent variables were significant. Wortman et al have speculated that a predictive use of SAINT in future may be the evaluation of alternate system configurations and operational procedures. They acknowledge one limitation, however, which is that the present model has only been applied to one type of mission and to one type of operator team.

Miller et al (1978) were also concerned with the modelling of remotely piloted vehicle control performance, for which an almost identical simulation facility existed. No predictions have as yet been compared with empirical data. Instead, the authors have challenged the construct validity of their model on logical grounds, and have assumed predictive validity will also be affected. In particular, they claim that

their model neglects cognitive processes, team performance and communication aspects of the task. It is difficult to know whether these inadequacies result from lack of refinement of the model or from some inherent shortcoming.

Generally, SAINT has wide application due to its flexibility. However, this is often achieved at the expense of using many free parameters. An associated problem is that, if users are free to develop their own sub-models, then there is less probability that a bank of standard parameter values will be available for the estimation process.

4.2.4.2 Design applications

From a logical perspective, SAINT should have a number of valuable pragmatic qualities. Both operator numbers and training levels may be model inputs. As such, SAINT should be well suited for predicting personnel requirements from conceptual system designs. Unfortunately, no application studies appear to have been conducted in which personnel-related issues required forecasting.

The effort that is commonly necessary to implement SAINT is probably greater than for any other network model, due to the large amounts of input data and the detailed analyses required. The use of network diagrams (Davis, 1982; Wohl et al, 1983) may alleviate this problem to some extent. Similarly, SAINT may have little applicability at conceptual stages of design if data are unavailable. In fact, no example has yet been found in which modelling was carried out in the absence of performance data gained from a prototype of the system. As discussed previously, modelling in some ways may then be superfluous for design purposes. On the other hand, the level of performance detail of SAINT is not fixed. As with the Siegel and Wolf model, if the analyst has the ability to conceive of a legitimate task network at a molar level of abstraction, then SAINT may have an early design application.

The facility for wedding human performance to system performance in SAINT is laudable from a system design viewpoint. It appears that the equations of motion of the system, for example, may be systematically varied as inputs to the model in order to observe their effects upon overall performance. The detail of the model output is also useful for similar reasons. If, for example, particular tasks are shown to have been repeated more often than was expected during the simulation exercise, further investigation may uncover the reasons and suggest system re-design. Whether the nature of SAINT is such that human-computer relationships may be modelled is an open question. It is encouraging that the flexibility of SAINT would allow informational parameters to be included in the model, provided that the analyst has sufficient expertise. Once again, the lack of application studies makes these issues difficult to resolve.

4.2.5 Summary

The strengths and weaknesses of the various network models have been continuously implied throughout this review. In summary, network models as a whole have some recognisable limitations (Pew et al, 1977):

(a) They are data-limited. Values for all input parameters must be derived by some means; either through estimation or identification.

(b) Task interactions are a constant problem. Whilst simulation methods do not incorporate combinatorial statistics, they still do not avoid the fact that the performance of an isolated task may not be the same as when it is embedded in a sequence of tasks. *Very often, practice and fatigue effects should at least be presumed.* One exception to this criticism is the SAINT methodology, which allows task performance to be moderated as a function of time or other variables.

(c) Cognitive processes are difficult to model, i.e. the models fail to capture performance in situations of 'low task density'. This is a somewhat contentious issue. In principle, there is no reason why a decision-making task, for instance, cannot be *characterised by the twin parameters of speed and accuracy.* On the other hand, the speed of decisions may not always be crucial. In addition, it may be inappropriate to rate decisions as only correct or incorrect. A continuous scale based on decision quality may be more useful (Alberts, 1980).

Perhaps paradoxically, network models may be said to have both wide and narrow domains of application. The network approach itself is generally applicable, as long as a system may be analysed into discrete tasks. *For each system, however, a new analysis must be performed in* order to formulate the model. Any one application, therefore, has a narrow focus. Considerable effort may have to be invested to perform this analysis and estimate the parameters. Accordingly, a major value of the modelling effort may lie in the insights gained through the analytic process itself rather than through the actual synthesis of performance predictions.

4.3 Information Processing Models

The most popular class of human models, information processing models, exist in many forms in the literature. Most are used to describe performance during a particular experimental paradigm, e.g. Sender's (1964) model of visual sampling behaviour. These models have the advantage of being detailed and well-validated for the situations which they describe. On the negative side, their relevance to C^2 is limited due to the fact that they are concerned with a relatively small portion of behaviour and tend to be very task-dependent. Models of human problem-solving performance are a case in point. In fact, *only one comprehensive* information processing model exists: the Human Operator Simulator (HOS) (Lane, Strieb, Glenn & Wherry, 1981).

4.3.1 The Human Operator Simulator (HOS)

HOS is actually an aggregation of human performance sub-models that have been derived from a literature survey and previous experimental work. The models are concerned with five areas of behaviour which, in sum, are thought sufficient to capture operator performance in complex systems. The models are all analytic in nature, i.e. given certain conditions, they compute operator performance via an equation. Duration of performance is the only dependent variable; the concept of error at

the sub-model level is neglected. (As regards overall output, however, error may still be defined as failure to complete a task within certain time limits).

The methodology of HOS has many parallels with that of network modelling. Model predictions are essentially derived by a synthetic process; that is, performance at the molar level is seen as the sum of performance on more molecular activities. In the terminology of Pew et al (1977), both HOS and network models reflect a 'bottom-up' approach to modelling. On the other hand, HOS requires qualitatively different input data and pre-modelling analysis. The analysis of 'basic procedures' of the proposed system takes precedence over analysis of the system into discrete tasks (Wherry, 1976). These procedures, together with system requirements, are used to form an individual program which then 'drives' the human performance models (Glenn, Zaklad & Wherry, 1982). The estimation of task-related parameters is dispensed with, because HOS draws from a bank of data which contains standard performance times relating to its primitive functions (Lane, Strieb & Leyland, 1980).

The philosophy underlying HOS is to construct a task-independent model that consequently has wide application (Wherry, 1976). This goal is said to be achieved by utilising performance sub-models which are extremely molecular in character. It is a premise of HOS that all operator actions are actually composed of relatively few basic activities, namely, information recall, mental computation, decision making, anatomy movement, control manipulations and relaxing. Despite its title as an 'information processing' model, therefore, manual actions are incorporated. HOS should not be seen so much as a new model as a new way of using existing models.

In some ways, HOS is prescriptive in nature. The sequence and types of operator actions are necessarily economical, i.e. it is presumed that operators do not perform superfluous actions. Parallel performance is also not possible, based on a single processing channel assumption. The lack of a specific facility for reliability data is justified by the claim that trained operators make no errors, at least within the performance domain of the sub-models (Lane, Strieb, Glenn & Wherry, 1981). However, errors may be programmed into the model if one desires, or untimely task performance may be defined as an error. Also, the outputs of the sub-models are almost exclusively deterministic, with the result that human variability cannot be represented at that level. The sub-model relating to memory is the only one which contains a stochastic process, so a distribution of performance times may be possible in that domain.

HOS contains parameters for individual differences in skill but not for team interactions. A single operator only may be modelled, and the performance of a group is derived by simple amalgamation. Communications variables are therefore difficult to model. HOS also presumes a stationary operator, that may be unrealistic in some circumstances.

In addition to the human performance models, HOS contains various other interacting programs. As mentioned already, procedural information forms a distinct component. Physical details of the system form another, including hardware details and workspace layouts. A third module allows a choice of methods of analysing the output data. HOS basically provides an

output of the various behaviours along a time-line. Statistics of the frequency of use of various items, or of task repetitions, etc, may then be compiled.

4.3.1.1 Validation studies

Explicit details of the validation of HOS are scarce. It appears that an incremental strategy towards validation is being adopted by the developers of HOS. That is, previously validated human performance sub-models have been selected and amalgamated. The predictions of the composite model have then been evaluated against performance data for reasonably simple, molecular tasks, such as human interaction with controls and displays. It is anticipated that HOS will be applied to the modelling of successively more complex behaviour in future.

Wherry (1976) cited a number of studies in which HOS was used to forecast performance for some low-level tasks. He claims, for example, that HOS made accurate predictions concerning reach times for certain sets of controls and reading times for certain displays. No details were given of the studies, so it is difficult to assess how great a necessity there was for estimation of free parameters. Construct validation has also been attempted, e.g. it was shown in a dial monitoring situation that the more variable sources of information received greater viewing time, as predicted.

Logically, HOS should be most valid in situations that are both well-defined and reasonably simple. In respect of the former point, the greater the potential for operators to deviate from an optimum sequence of actions, the less accurate the predictions of HOS will be. Secondly, the inaccuracies of HOS are likely to be compounded when the predictions from simple tasks are aggregated into more complex tasks. There is even some evidence that the basic human performance sub-models of HOS may not generalise to all situations. For example, Glenn et al (1982) believe that the models that are concerned with the motor aspects of behaviour are most valid (and general), whilst the decision-making component of HOS may be in need of refinement.

As for the oft-repeated claim that HOS is the most generalisable and comprehensive of all models, both positive and negative support exists. The use of elementary human performance sub-models ensures task-independence, as those models may be applied to any situation with little modification. On the other hand, those models are driven by a procedural description of the system, and the means of achieving this description appears to be treated somewhat mysteriously. In practice, the procedural description may well be equivalent to an analysis of the system into discrete tasks. For every system, therefore, a new analysis may be necessary. As with network models, the procedural approach of HOS may be generally applicable whereas particular realisations of HOS may be system-dependent.

Possibly the most ambitious application of HOS so far has been the modelling of the crew station onboard a U.S. Navy anti-submarine aircraft (Lane, Strieb & Leyland, 1980). Experience had shown that the installation of a Forward Looking Infra Red (FLIR) surveillance system had actually caused a decrement in performance, due to operators being overloaded with simultaneous navigation, detection and interception tasks. A scenario was devised in which the crew had to maximise the

gathering of intelligence whilst minimising flight transit time. The scenario formed the basis for the HOS modelling, and simulation did suggest that the addition of the FLIR system was counter-productive to intelligence gathering, in comparison to a standard system. This demonstration is somewhat less impressive due to the fact that the simulation was validated retrospectively, i.e. an operational system was available for analysis before modelling, and estimation of parameters in advance was not necessary.

4.3.1.2 Design applications

The work of Lane, Strieb and Leyland (1980) also provides information on a design application of HOS. Following the validation phase, in which HOS captured the deleterious effects of the FLIR system, the performance of the system under an automated navigation capability was simulated. The results suggested that this new system would be beneficial to performance through reduction of operator workload. Thus, although the validation of the model may be criticised for being retrospective in nature, the simulation was still productive in that it enabled an evaluation to be made of the effects of a re-allocation of function within the system.

Unusually, an attempt was also made to predict the cost of operation of these various systems, as well as their efficiencies. Cost was mainly hardware-related (e.g. fuel usage) but did include personnel working time. The model was not sensitive to the differences between the systems, largely because all were predicted to require the same transit time for fulfillment of the mission.

As yet, the predicted efficacy of the automated tracking system is hypothetical and has not been tested in practice. The investigators, in addition, have speculated that they will be able to predict the effects of different tactical strategies within the system, through modification of the procedural component of the model. Apart from the predictions of HOS per se, Lane et al have also stressed the value of their preliminary systems analysis, which, for example, allowed them to identify approximately 100 key decision-points.

An analysis of the construction of HOS allows further comments to be made regarding the model's pragmatic qualities. Generally, the treatment of personnel issues is not extensive. The model contains parameters for individual differences in elemental abilities such as sensing and remembering (Wherry, 1976), but the relation to overall training requirements is uncertain. According to Meister (1971b), the outputs of the model may indirectly suggest what level of operator skill is required (namely, through the identification of those tasks which are performed in an untimely fashion). As the model does not accommodate group interactions, personnel numbers may only be inferred from the performances of individuals.

HOS is probably the most sophisticated human performance model for suggesting design solutions, for a number of reasons. First, the model should be quite sensitive to design parameters, as both hardware details and operational procedures are specifically required as inputs and form separate program modules. Workspace layout must be particuarly well-defined. Secondly, the statistical detail of HOS's output allows a comprehensive check to be made on the functioning of the hypothetical

system. The usage of hardware items and even body parts may be tabulated. Link analysis may be performed to assess the spatial aspects of a working environment. Situations of high workload may be identified by noting which task completion times exceed the allowable constraints. High memory loads may be identified through a tabulation of use of the memory sub-model.

Given that HOS integrates human and systemic variables into the one model, it would be useful to assess its relevance for modelling human-computer relationships. Unfortunately, HOS has not been applied towards such a goal. Lane, Strieb and Leyland (1980) claimed that software based devices may be incorporated into the model, which is encouraging, although no further details were given. They also speculated that the efficacy of decision-aiding systems may be evaluated through HOS in future.

HOS has probably the greatest potential of all bottom-up models for cognitive modelling and subsequent engineering. If these considerations arise relatively late in the development cycle (such as when deciding the content of software dialogue, for example), then HOS may be a useful evaluative tool. Decisions regarding the allocation of function between humans and computers (in a gross sense), however, may have to be made by some other means.

The degree of effort required to implement HOS (in comparison to network models) is difficult to gauge. Somewhat paradoxically, it has been claimed that the use of HOS may be impractical during early stages of system design due to large data requirements (Meister, 1971b), whilst an advantage of HOS has been claimed to be that it dispenses with the necessity of supplying parameter values for all tasks (Lane, Strieb & Leyland, 1980). On balance, the former view is probably closer to the truth. HOS has a requirement for translation of system goals into a set of operator procedures. The fact that a motivated and well-trained operator is presumed, means that the modelling of non-essential activities may be avoided; but, generally, the amount of pre-modelling analytic effort that is involved will be the same for the HOS and network methodologies. That is, most task networks are normative in character. HOS then computes task times automatically by invoking the elemental performance sub-models, that avoids the data limitation problems of network models.

On the other hand, the applicability of HOS at conceptual stages of design may be comparatively low, due to other input requirements. The details of workspace layout, for example, must be supplied, which may be difficult at that time. That is, HOS appears to lack the facility for being implemented at different levels of abstraction. In its present form, it is specialised for interface design (Wherry, 1976) and consequently, requires a system description which is normally available at that stage of development. Even if the required details are available for a conceptual system, Pew et al (1977) suggest that tne predictions of HOS may be obtained by a process that is unnecessarily elaborate in comparison to a more 'global' means of evaluation.

4.4 Control Theoretic Models

The 'bottom-up' approach to human performance modelling, as exemplified by network models and HOS, is essentially a psychological endeavour. The elementary tasks of both models (from which overall

performance is synthesised) are most frequently described in terms of the activities of an operator. By contrast, the class of control-theoretic models attempt to describe the behaviour of the human only insofar as is necessary for the achievement of system goals. As a consequence, it is convenient to represent the activities of the human in machine-like terms (Rouse, 1980), i.e. an engineering-style description predominates.

As implied by the title, a central concept of control theory is maintenance of the system in a particular state, as defined by one or more variables. The system may initially be in an undesirable state, or may deviate from a given state due to influences which are either external or internal to the system. It is presumed that the human, as a controller, has the ability to sense the state of the system and may take actions to maintain or restore the state to a given value. It is also presumed that the human is situated in a closed-loop, i.e. feedback allows the possibility of error correction.

The most frequent application of control theory has been to manual control behaviour, and system state has most commonly been defined in terms of position. Human tracking performance (either pursuit or compensatory) is a classic example. In addition, higher-order system states may require control, such as when an operator must adjust the variables of velocity and acceleration of a vehicle. Further, increasing emphasis is being placed on the decisions (rather than the ph···ical actions) of the operator as a means of achieving system control (Pew & Baron, 1982).

All control theoretic models share a common methodology. The prime requirement is for a system which may be described (mathematically) in terms of control of some state variables in a disturbing environment. Thus, various system goals must be defined, along with a statement of the system dynamics. Under a presumption that the system will attain stability (or some other state), human performance is predicted (analytically) as that which closes the feedback loop of the system by providing control inputs.

Two significance consequences follow from this methodology. The first is that behaviour is prescribed, i.e. the activities of the human are presumed to be optimal (within certain limits) with respect to system goals, with the over-riding goal being minimisation of system deviation from a defined state. The second consequence is that it is only those characteristics of behaviour that are sufficient for achieving system goals that are prescribed. Human processes as represented by control theoretic models are superficially isomorphic to actual human processes, but no more. In other words, the purpose of control theoretic modelling is such that the degree of construct validity that is required is relatively low.

These features of control theory constitute a 'top-down' approach to performance prediction (Pew et al, 1977). An alternate, but closely related, methodology is to commence with a specific behavioural description (although one which is compatible with the control theory framework) and a statement of system dynamics. System performance may then be predicted, with a familiar output being a plot of system error as a function of time. If allowable limits may be specified for that error, then the probability of the systems exceeding those limits may also be predicted.

Two types of predictions of control-oriented models may thus be tested against empirical data, with one prediction relating to human performance and the other to system performance. Validation of the human performance aspects of a control model is a prerequisite for prediction at the systems level. As it is presumed that an accurate model of system dynamics is always available, precise system predictions (such as tracking error) are frequently regarded as evidence of the validity of the human performance model.

Various control models differ in the assumptions that they make about human characteristics. Firstly, the human is always assumed to be constrained in comparison to an ideal controller (i.e. one which is both instantaneous and error-free). The nature of these limitations is one distinguishing feature of the models. Secondly, different models presume different relationships between the inputs and the outputs of the operator. As an example, Pew et al (1977) distinguished three broad categories, namely, quasi-linear control models, including the crossover model (McRuer, Graham, Drendel & Reisener; 1965), fixed-form parameter optimisation models, including paper pilot (Anderson, 1970) and the optimal control model (Kleinman, Baron & Levison, 1970).

All have been validated in certain situations, but their range of application differs. In particular, tasks that are more complex than the simple uni-dimensional tracking task require enrichment of the human performance model. The optimal control model and its derivatives represent the most advanced state-of-the-art in modelling human performance via control theory, and have the greatest comprehensiveness. For this reason, further details of the optimal control model alone will be given.

The optimal control model (Kleinman, Baron & Levison, 1970) as a description of human behaviour divides logically into four parts. Sensory capabilities are represented in the perceptual component, as it is assumed that the limitations of the human result in a perception of system output which is both delayed and noisy. Motor aspects of behaviour are also assumed to be delayed (due to neuromuscular lag) and noisy. Intervening between stimulus and response is the central processing component of the model, that estimates the state of the system in an optimal fashion, i.e. this function compensates in part for the noisy perceptual data. An optimal predictor is also incorporated which offsets the effects of delay in the perceptual component. Lastly, the cost function component of the model describes the human's control strategy. That is, whilst the operator is presumed to minimise overall system error, it is possible to specify relative penalties for various system errors and degrees of control 'effort'.

From a model user's point of view, both the observation and motor delays require parameter estimation. Similarly, the noisy quality of both perception and motor output is represented by two signal-noise ratio parameters, respectively. The values of these parameters are considered to be task independent, and empirical studies have reinforced this view (Kleinman et al, 1970). That is, human performance under a range of conditions and in different systems may be adequately represented in control theory terms using constant values of these parameters. The parameters which specify the cost function are presumed to change in different systems and must be estimated by any available means before system performance may be predicted.

Conversely, given a statement of system dynamics and a presumption of a particular control strategy, the human's control behaviour may be predicted. In that case, a describing function is yielded that contains two parameters which are good indices of the human's performance (Kleinman et al, 1970). However, whilst the optimal control model may be used to predict both system and human performance, it is more useful for explanation rather than prediction of human performance data. In short, the focus of the analysis is on human performance as a component of the system, rather than as an isolated phenomenon.

From a designer's viewpoint, the crucial aspect of modelling via control theory is the statement of system dynamics. As for the impact on human performance, the features of different system designs are usually related to the perceptual and motor components of the optimal control model. For example, the amount of information in a display may reasonably be expected to influence the operator's perceptual capabilities, and if this feature may be represented mathematically, then the effects on system performance may be predicted. Similarly, the effects of different relationships between the human's control inputs and the system output may be investigated.

In practice, the operator in a complex system is faced with multiple displays. The optimal control model accommodates this situation by incorporating an optimal scanning model of performance (Baron, Kleinman & Levison, 1970). The effects of attention allocation strategies may be studied by assigning suitable weighting parameter values to the scanning of each display. (In the absence of a specific theory, a priori assignment of attention allocation may very well be difficult). Possibly the most advanced derivative of optimal control theory in this context is DEMON (Pew & Baron, 1982) which models the multi-task control of remotely piloted vehicles. In addition, the optimal control model has been modified to cope with decision making performance. The Dynamic Decision Model (Pew & Baron, 1982) dispenses with both the motor component and cost function, and in their place substitutes an optimal decision-cost matrix that minimises decision error.

4.4.1 Validation studies

The earliest validation studies of the optimal control model were concerned with demonstrating that the values of the parameters that represent human perceptual/motor limitations may be held constant under a variety of conditions, whilst system performance (such as tracking error) may be accurately predicted. Thus, validation was indirectly obtained for the human performance structure of the model, in addition to demonstrating the task independence of its parameters. Later studies have shown more concern for the constructs of the model by comparing its prescriptions of human performance with actual behavioural data. Most extensions of the model to different tasks have required some refinement of the structure of the model, so more attention has been directed towards its construct validity. For example, Kleinman, Pattipati and Ephrath (1980) found it necessary to postulate a short-term memory component in order to account for tracking behaviour during periods when the target was blanked out.

As for military applications, the model has been validated for a range of conditions, including pilot hovering tasks (Baron et al, 1970: Johannsen & Govindaraj, 1980) and tracking performance of an anti-aircraft

gunner (Kleinman et al, 1980). Performance data from simulators has been utilised more often than that from operational systems. At a more behavioural level, predictions regarding scanning behaviour and attentional allocation have been shown to be accurate (Baron et al, 1970; Johannsen & Govindaraj, 1980).

Within its domain of application, great generality is claimed for the optimal control model (Pew et al, 1977). On the negative side, a frequent criticism is that the optimal control model is over-determined, i.e. it contains more parameters than are necessary to describe the input-output behaviour of the human. A consequence is that no unique parameter set accounts for system performance data. One solution which has been suggested is to simplify the model (Phatak, Weinert & Segall, 1975) and thus reduce the number of parameters. In particular, it may be reasonable to presume no perceptual delay (if system state is easily predicted) and no motor noise (for the well-trained operator). Levison (1982) believes that the use of standard rules for adjusting parameter values may compensate in part for over-determination of the model.

4.4.2 Design applications

A range of design applications of the control theoretic models exist in the literature. Unfortunately, many must remain hypothetical, for the operational systems have not been built. The studies also show a strong control/display bias which suggests that the models, despite the claim for wide generality, may be limited in application. On the other hand, Baron and Levison (1977) and Curry, Kleinman and Hoffman (1977) have stressed that the minute details of display design, such as legibility, are not necessarily addressed by the optimal control model; rather, informational aspects are paramount.

With respect to display design, two factors have generally been investigated. The first is display arrangement; for example, Bhattacharyya, Prasad and Sarma (1972) used the optimal control model to predict the best arrangement for a jet transport landing task, the conclusion being that the attitude indicator should be central in order to avoid excess scanning by the pilot. More commonly, the influence of various display variables has been investigated, often by predicting the system performance which results from a display which has been augmented in some way over a standard display, e.g., Baron and Levison (1977). Augmentation most often consists of the addition of predictive and/or higher order information regarding some of the display variables, and both factors have been shown to be significant, at least in a NASA hovering task (Johannsen & Govindaraj, 1980).

The effects of displays on operator behaviour have also been modelled as part of the evaluation process. For example, Baron and Levison (1975) analysed the effects of a hypothetical advanced display in a NASA vertical take-off and landing task. They concluded that the display would induce changes in the pilot's attention allocation strategy and would also result in changes in the ratio of state estimation to control behaviour time. Connelly (1977) predicted and subsequently demonstrated that an advanced display increased the preview range and, therefore, the information processing capacity of a naval deck officer.

In one way, the conclusion regarding augmented displays is pre-determined by the structure of the optimal control model, for the operator

is presumed to receive feedback for all system variables that are being controlled. It is reasonable that the more detailed the feedback, the better the control which is exerted, and this is assumed to some extent (although additional display variables must also compete for attention). Similarly, Hess (1981) showed that an automated flight direction system in a helicopter could be expected to improve system performance. Once again, it is reasonable that the automation of one control loop will release resources that may be applied to other loops. As regards the assessment of such advanced systems, the value of the model probably lies not so much in predicting an improvement in system performance as in delineating the extent of that improvement. A secondary benefit is that the model also produces a mathematical description of the associated change in human performance.

As already discussed, control theory regards human and system performance as inextricable. A consequence is that alternative designs may readily be evaluated and compared. As for design solutions, there is a necessity for the modeller to be able to interpret the effects which system dynamics have on human performance. With the optimal control model at least, the fact that performance is divided into logical components probably assists the solution process. In particular, the effects of feedback characteristics, form of system disturbance, control gain, etc, may readily be investigated.

As regards the evaluation of systems in which man and computer interact, few specific application studies have been found. In principle at least, the effects of a decision-aiding system, for instance, could be modelled by suitable modifications of either the perceptual or information processing components of the human performance model. Alternatively, a separate artificial intelligence component could be inserted. Other applications undoubtedly exist. Two considerations presently suggest that modelling human-computer systems in control theory terms may be inappropriate. One is that the modelling of non-continuous, non-manual processes can rarely be achieved, although some workers (Pew et al, 1977; Pew & Baron, 1982) believe that this difficulty may be overcome in principle. The second, and related, issue is that the extension of control theory to other than control/display design is contentious.

The effort and expertise that is necessary to implement the optimal control model is qualitatively different from that required for network modelling. It is a characteristic of the top-down approach that detailed pre-modelling task analysis of the activities of the operator is not essential. (As discussed, HOS also shares this characteristic although, in practice, the analysis must still be carried to a fairly detailed level.) In addition, extensive estimation of parameters which describe human performance is overcome both because tasks are not individualized and because most parameter values are regarded as system-independent. In place of the network programme, control theory requires a mathematical representation of the system dynamics and goals, the difficulty of which depends on the modeller's expertise (naturally) and the extent to which the system conforms to a control theoretic framework. As regards the optimal control model, the specification of a cost function (that described the human's control strategy) may also be a 'non-trivial' requirement (Kleinman et al, 1970).

The suitability of control theoretic modelling at conceptual stages of design is probably low. The requirement for a mathematical

statement of system dynamics implies that certain hardware details have become fixed, although Curry et al (1977) have illustrated how the model may be implemented at different levels of abstraction. On the other hand, control modelling is probably most useful for evaluating re-design, e.g., automation, of an operational system, and the application studies in the literature support this view.

From a pragmatic basis, one of the most serious deficiences of control modelling is the neglect of both individual differences and team performance. The neglect of differential ability is a direct function of the top-down approach, i.e., as behaviour is prescribed, it is presumed that all operators will conform to this requirement. Sub-optimal performance invalidates application studies of the model, e.g., Connelly (1977), Johannsen and Govindaraj (1980). Training issues therefore cannot be easily resolved. Similarly, control theoretic models in their present form do not incorporate group behaviour. The numbers of personnel required to operate a system must be derived from individual performances, thus neglecting team interactions.

4.5 Queuing Models

Queuing theory represents another top-down approach to systems analysis. The models derived from this approach thus have the basic qualities of prescription of human performance (bounded by certain constraints) in sufficient (i.e., engineering) terms. In contrast to control theory, queuing theory does not characterise the deviation of system state from some defined value. It is more applicable to systems in which a number of operations compete for processing, an assumption being that only one may be performed at once. The human is presumed to process operations by a strategy that minimises a (pre-defined) cost to the system of not performing those operations. The human's role in such systems is, using the terminology of Rouse (1980), a 'time sharer' or 'attention allocator' rather than an 'error nullifier', as in control systems.

Queuing theory basically requires an analysis of system dynamics in terms of the 'traffic' characteristics of the multiple tasks. Depending on one's formulation, the number of tasks, their arrival rate and their distribution in time are important independent variables. It is possible that these values may change with time. Task priorities must be specified in order that the cost of neglecting tasks may be represented mathematically. The limitations of the human must also be defined, largely in terms of the operator's performance as a 'server'. In particular, service rate is an important variable. All these variables, and more, constitute the independent parameters of the model, and require estimation. Unfortunately, it is not yet possible to describe human performance independently of the system context in which it occurs, which reduces the generality of the model.

The outputs of queuing models comprise both system and human performance. At a certain time, it is possible to predict the number of tasks completed and the average waiting time, for example. Assuming that the conditions of the system are such that the operator will have some idle time, the average degree of server occupancy may also be predicted. The magnitude of this variable may be interpreted as an index of operator workload (Rouse, 1980), and may be expressed as a function of task type. Queuing theory is therefore more concerned with how the human

co-ordinates a set of tasks rather than with the performance of any one task, in contrast to control theory (Rouse, 1980).

In order to formulate queuing models, the system must satisfy a number of fairly rigorous mathematical conditions. For example, difficulties are encountered if tasks do not arrive independently of each other (Rouse, 1980), because an analytic solution of the model equations may not be possible. Lack of a theoretical data base also means that many parameters must be left free to vary. Prediction is then weakened by retrospective fitting of the model to empirical data. Taken together, these considerations imply that the fundamental requirements for both validity and generality of queuing models are rarely satisfied. The most successful queuing models have been applied to very well-defined situations and have consequently tended to neglect performance at a systems level, e.g., the visual sampling model of Carbonell, Ward and Senders (1968). But for recent developments, queuing theoretic models would have been relegated in this report to the category of those inapplicable to system design and procurement, such as many information processing and problem-solving models.

Chu and Rouse (1979), however, have applied a queuing model to prediction of pilot performance in a multi-task situation. Basically, activities necessary to control the aircraft were modelled as a subset of the total mission, although these activities were regarded as primary. A secondary group of activities included monitoring and resetting of dials that yielded information on the physical state of the plane. A control model was actually combined with the queuing model in order to predict control and subsystem performance, respectively. That is, it was possible to account for aircraft trajectory simultaneously with average waiting times, etc., for the secondary tasks. Further, pilot workload (i.e., server engagement) for the secondary tasks was a model output. All outputs of the combined model have been tested against data from a simulator, although parameter estimation is required from the same data set.

The fact that the queuing model of Chu and Rouse (1979) has been shown to be quantitative and valid at a systems level is significant. It allowed an extension of the top-down modelling approach to a situation of multiple-task performances, not all of which involved manual control. Secondly, Rouse (1980) considers that queuing formulations may be the most appropriate means of representing human-computer interaction, provided that the necessary assumptions can be met. In particular, queuing theory has the flexibility to incorporate multiple servers, not all of whom must be human. It is theoretically possible to insert a model of a machine which has its own serving characteristics and thence predict overall system performance.

As a speculative example that is relevant to C^2, a combined human-computer queuing model could be applied to a situation of processing multiple sources of information. Given a statement of both system dynamics and human limitations, it might then be possible to specify the server characteristics of a computer that would be necessary to achieve system goals. Alternatively, given the server characteristics of the computer, it might then be possible to predict the necessary engagement performance of the operator(s).

Unfortunately, no such design applications of queuing theory yet exist. Chu and Rouse (1979) did apply their model to the evaluation of a pilot aiding system, but from a different perspective. In that instance, the aiding device had access to a real-time queuing model of system state, based on performance of the secondary activities. Broadly, if the computer 'sensed' that secondary task performance had become inadequate (due to the occurrence of a number of unserviced tasks), it was able to assume automatic control of the navigation system. The functions of the pilot therefore changed dynamically in an attempt to maintain workload at a constant level. Data from the prototype experiments did suggest that the aiding system could result in better flight control, better subsystem performance and more constant pilot workload.

Similarly, Chu, Chen, Clark and Freedy (1982) evaluated a pilot decision-aid that was based on a queuing theoretic model of system performance. The aid automatically selected information for the pilot's attention, and presentation of this information depended on subtask performance. A system prototype was used to demonstrate that the aid could improve flight control under these conditions.

Whilst this latter study and that of Chu and Rouse (1979) were not oriented towards assessing the use of queuing theoretic models as a design tool, certain implications may be drawn. Generally, if it is possible to capture the performance of an operational system (or its prototype) in terms of a queuing model, then the methodology exists for predicting system and operator performance from certain input conditions. Whether this end is realised depends largely on whether the number of free parameters in queuing models can be reduced with further refinement.

The potential of queuing theoretic models for C^2 system design may be evaluated on logical grounds. As with control theory, it may be said that the top-down approach facilitates design solutions, due to the prescriptive relationship between system dynamics and goals and human performance. As queuing theory is more concerned with task co-ordination than with task performance per se, it could be said that the model is less directly sensitive to hardware variables and more so to procedural variables. As discussed, human-computer interaction in well-defined circumstances may be amenable to modelling via queuing theory. Decision aiding may be evaluated. The applicability of the technique at conceptual stages of design is difficult to judge from the literature, but is probably comparable to the relatively low applicability of control theory.

Queuing models treat personnel parameters in a more comprehensive fashion than other top-down approaches. Whilst constrained optimality of human behaviour is assumed, a parameter exists that could be said to characterise individual differences in skill. This is the 'server rate' variable, although it may be difficult to relate to skill or training level. In addition, the models may incorporate multiple servers (Rouse, 1980) which would allow determination of the numbers of operators required from a design. Team interactions are not addressed directly, although the effects of a heterogeneous population of servers, possibly acting with different task priorities, may be modelled.

4.6 Composite models

The use of composite models has frequently been advocated as a means of extending the domain of system and human performance prediction, e.g., Sheridan and Ferrell (1974), Rouse (1980). The optimal control model is actually such an example, because it amalgamates concepts derived from both estimation and control theory (Rouse, 1980). Combinations of control models and queuing formulations are also possible, e.g., Walden and Rouse (1978).

A more significant combination may be that of two or more models that reflect totally different philosophies towards modelling. This appears to be the major concern of workers in this field. Logically, if the inherent deficiencies of one model are precisely the strengths of another, then a synthesis may be desirable (Pew and Baron, 1982). A wedding of the top-down and bottom-up approaches is claimed to be especially productive in this context.

The motivation for such composite modelling appears to stem particularly from the inability of control theoretic modelling readily to incorporate monitoring and supervisory behaviour. The optimal control model has the virtue of yielding quantitative predictions at a general system level, but it loses applicability in some situations of discrete or intermittent performance. This difficulty has been regarded as potentially soluble with further refinement of the model (Pew et al, 1977) while others consider that the modelling of non-continuous behaviour by control-theory is unsound in principle (Beishon, 1967; Rouse, 1980). In any event, the use of control theory with some form of bottom-up approach provides a powerful and general means of systems modelling.

4.6.1 PROCRU

The best known composite model in this regard is PROCRU (Procedure Oriented Crew Model) (Baron, Muralidharan, Lancraft and Zacharias, 1980). The model has been developed for NASA in order to analyse/predict crew and system behaviour during an aircraft approach to landing. Unfortunately, no validation or design studies have yet been reported, but sensitivity analyses have ensured the model's internal consistency.

The formulation of PROCRU commences with a systems analysis. Basic procedures and their sequences must be identified by studying similar systems, reviewing training manuals, etc., in an analogous fashion to the requirements for network modelling. At the same time, an analysis is made of system goals and dynamics in order that the mission may be given a control-theoretic interpretation. Mission performance is not simply predicted by synthesising the parameters that describe the performance of each procedure; rather, the choice of whether to execute any procedure is based on the assumption that the crew will maximise net gain to the mission. The choice and performance of any procedure depends critically on the state of the system at the time, which in turn depends on previous activities. For example, if the simulation dictates that the trajectory of the aircraft is beyond pre-defined bounds, then control activity may be given precedence over other activities. Additionally, the constraints for that activity may also alter as a function of the state of the system. Human performance is thus modelled in a far more dynamic fashion than occurs in network modelling.

It may be said that the 'trigger' for all actions within PROCRU has two sources (Baron et al, 1980). One is the prescriptions of the control theoretic model, which are a function of system state. In addition, events which are 'external' to the model may be programmed, that is a heritage of the bottom-up component. For example, communication activities, as identified in the preliminary systems analysis, may be inserted and generally consume resources that would otherwise be available for monitoring/control. In this manner, the domain of application is greater than that of the standard top-down approach.

The prescriptive human performance components of PROCRU also appear to be more refined than those of the optimal control model. Perceptual activity includes audition as well as vision, for example. The information processing sub-model includes a facility for memory processes. As with other control models, however, only individual performance is predicted. Team interactions are not prescribed, but may be simulated indirectly through the procedural network if the modeller has the necessary expertise.

The basic dimension of performance in PROCRU is time. The outputs of the model include aircraft trajectory as a function of mission stage, and also a time-line of catalogued activities of the crew. This computation relies on two factors. First, the times of completion for monitoring and control activities are calculated within the conditions specified by the top-down model. Performance time for discrete activities, by contrast, must be supplied by an analyst. PROCRU does not consider performance error directly, i.e., task reliability is not an input. However, the execution of control or monitoring procedures may be beyond tolerable limits due to perceptual, informational or workload considerations.

4.6.2 Other models

A second example of a synthesis between the top-down and bottom-up approaches is provided by the work of Kraiss (1981). In that case, a combination of the cross-over model and network simulation via SAINT was carried out. Once again, the choice of whether to execute discrete or control theory related procedures was prescribed by the top-down methodology. As with PROCRU, the major outputs of the model are both system and operator performance as a function of time.

The model has been applied to the evaluation of two submarine displays; one being a standard and the other being augmented with predictive information concerning system variables. The author claims to have validated the model by comparing its predictions of system performance (in particular scenario) with those derived from human interaction with a simulator. However, it is unclear whether the prototype existed before or after the formulation of the model, which could detract from any claim for predictive validity. Despite this criticism, model outputs regarding errors of pitch, heading depth, etc., correlated well with data from the simulator.

The evaluative component of this study suggested that the predictor display yielded superior performance at both a systems and an individual level. All measures of the deviation of system state from its prescribed value were predicted to be smaller when the augmented display

was employed. As the human's attentional allocation strategy was prescribed by the top-down model, it was also possible to make predictions regarding human performance under the two conditions. Specifically, it was shown that the augmented condition could be expected to be associated with a more regular scanning of the display and also more idle time. This result was interpreted as suggesting lower operator workload for that display.

4.6.3 Prescriptive aspects of bottom-up models

It should be noted that, whilst composite models are unique in combining two divergent approaches towards systems prediction, some bottom-up models do contain elements of the top-down approach. Both SAINT and HOS are top-down in the sense that system requirements determine their task sequences to some extent. That is, while the tasks network is determined before the modelling in a static fashion, although incorporating precedence relationships, the moment-to-moment state of the system may also influence that network. Human performance is thus modelled as a dynamic function of system requirements.

Generally, Pew and Baron (1982) believe that the most successful models have been constructed for situations in which the behaviour of the operator is highly constrained by the demands of the system. Two consequences for bottom-up modelling appear to follow. First, Kraiss (1981) has indicated that it is desirable for network models to be formulated from a normative task sequence. That is, it is preferable to devise the network by logical analysis of the system (possibly using a scenario) rather than by observing operator behaviour directly.

Secondly, Siegel et al (1977) have emphasised the importance of time constraints to ensure the predictive validity of their models. That is, the most successful models are those in which operator tasks are paced by the demands of the system, are crucial, and in which performance is sensitive to the necessity for task repetition. Many human functions within aviation systems fulfill these requirements reasonably well. As an aside, it is interesting that those models which forecast system maintainability have generally employed a different methodology; one in which task performance is estimated using data from human judges, e.g., Burger, Knowles and Wulfeck (1970). (An exception is the study of Proctor and Khalid (1971) in which a network model was used). One reason for this difference may be that the maintenance function is often driven by a demand for accuracy within generous time-constraints.

4.6.4 Prediction of emergent properties of systems

Within a modelling context, the emergent properties of a system may be regarded as those that cannot be forecast from data about the individual tasks which comprise that system. The emergent properties of a system are usually discovered by observing the functioning of the system as a whole rather than by observing (or analysing) individual elements of that system. This report has had an indirect concern for such matters when discussing the difficulty of accommodating task interactions within simple analytic models. That is, the performance of a task individually may not be identical to that when embedded in a task sequence because, for example, parallel rather than sequential performance may occur.

Bottom-up models suffer the general problem in that synthesis of performance data may not be an accurate representation of system functioning. Both Miller et al (1978) and Pew and Baron (1982) have made this criticism and have implied that composite modelling techniques may provide a solution. On the other hand, it is contentious (in a philosophical sense) whether any models may actually predict unforeseen events, or emergent properties. All models are constructed upon some abstraction of the system made by the modeller. Whilst models may contain some stochastic processes, it is doubtful whether any of these processes are effectively unanticipated.

It is probably more reasonable to claim that models differ in the extent to which human performance is represented dynamically rather than statistically. By this interpretation, both SAINT and HOS have relatively good predictive qualities. The use of precedence relationships and Monte Carlo simulation within network modelling may also be seen as attempts to mirror the variable conditions under which operational performance occurs.

A second sense in which one may speak about the emergent properties of systems is when considering the effects of operator strategies upon system performance (see Section 6.8). These strategies are difficult to anticipate, especially if a similar system is not available for study. Needless to say, no direct example has been found of models that may accommodate operator strategies. Most models tend to prescribe operator activity and are most valid when applied to systems which constrain that behaviour to a significant extent. That is, the greater the flexibility of a system, the less possible it is to predict system performance. Flexibility, however, does have definite advantages, which are discussed in Section 6.10.

4.7 Model Summary

A range of performance models have been evaluated along both theoretical and pragmatic dimensions. In most cases, the evaluation was performed without reference to empirical data because of a lack of such information. Some models have not reached a sufficient stage of development for a rigorous evaluation to be carried out. However it may be expected that the predictive validities of the models will increase with greater use.

Validity was indicated as a prerequisite for the inclusion of a model in this report. Given that this condition is always satisfied to some minimal extent, it could be said that the greatest emphasis of workers in the field of human performance modelling is on model generality. This consideration produces two closely related issues. The first is the degree to which the models are independent of the particular task or system. The second is the extent to which they address cognitive behaviour. Models derived from both the bottom-up and top-down approaches have been subjected to criticism from these perspectives.

As discussed, the methodology of bottom-up modelling may be applied to virtually any system, as some form of task network may usually be envisaged. On the other hand, suitable input data may be difficult to find and task interactions may reduce the accuracy of the model's predictions. The methodology of the top-down approach is in principle

less general, because the system must be sufficiently well defined for a mathematical statement of system dynamics and goals to be made. In addition, there must be reasons for believing that the behaviour of the human will be constrained enough to reflect system demand closely. If these conditions can be satisfied, then the top-down approach has the advantage of providing detailed, quantitative performance forecasts. Within the top-down approach, it appears that a control theoretic interpretation of system dynamics is the most widely applicable analogy (Rouse, 1980).

As for the task independence of models, there are two aspects to consider. First, no model can be applied to new systems without some form of modification. Systems analysis is invariably required, either to construct a task network (for the bottom-up approach) or to derive system goals and dynamics (for the top-down approach). The parameter estimation process appears to be the area in which task-independence becomes controversial. The most task-independent models are those whose parameters reflect basic human processes rather than characteristics of the system. If a bank of human performance data exists, then estimation is facilitated.

In the case of bottom-up modelling, independence of the models is achieved by formulating them at a low level, i.e., so that the tasks which constitute the network are a series of relatively molecular processes. HOS examplifies this approach. (A reciprocal dilemma then arises, namely, that task interactions become more prominent.) It could also be said that the independence of the top-down models is achieved by a similar strategy; that of formulating them at a more psychologically detailed level. The optimal control model is the most well developed example in this respect. The strategy is somewhat contrary to the top-down philosophy and probably results in increased model complexity.

No models have yet coped adequately with cognitive behaviour and have also generated quantitative predictions at a systems level. The most cognitive models, e.g., problem solving or information processing models, tend to be both task dependent and descriptive. (In one way, this result is not surprising because, as argued, there is frequently a trade-off between model generality and validity). In principle, there is no reason why the large number of algorithms and information flow diagrams which have been produced by workers in the cognitive field could not be integrated into a more comprehensive systems network model. In practice, task interactions and scarcity of input data are a perennial problem, as is the analytic expertise required. HOS overcomes the problem of input data to some extent, and probably shows the greatest potential of the bottom-up approach for modelling cognitive behaviour. By contrast, the top-down approach addresses cognition only insofar as is sufficient for achieving system goals, and in so doing yields quantitative, systemic predictions. Queuing theoretic models probably have the greatest potential in this regard, although further refinerent is needed.

4.7.1 Model comparisons

The preceding discussion of the generality, or domain of application, of models suggests that a comparative evaluation of models is difficult. Although all models in this report may be used to assess the human engineering adequacy of systems (or, more specifically, the operability of systems), the different techniques tend to answer different questions about the systems which they model.

In particular, the distinction between the psychological and engineering approach to modelling (or between the top-down and bottom-up approaches) is a source of difficulty. As discussed, the focus of top-down models is on the relationship between an optimal human performance and system performance. In a design sense, the most frequent methodology employed is to examine the effects of some innovation (such as an augmented display, or automation of one control loop) upon system performance. By contrast, the bottom-up approach has most frequently been used to assess the impact of new designs upon human performance. Following such an exercise it is then decided, on a priori grounds, whether the forecast performance will be sufficient to meet system demands or whether the required performance is unreasonable.

Both of these approaches have their merits, as has been demonstrated by the design studies reviewed in this report. From a design perspective, the top-down approach has the advantage of predicting system effectiveness, but this may only be done in situations in which the response of the human is well understood mathematically, such as when engaged in manual control or queue service. It is encouraging that bottom up models are beginning to incorporate system effectiveness parameters (namely, the later Siegel & Wolf models and SAINT), as those models should generate better design information. The composite modelling approach has specifically addressed this issue.

For the sake of model comparison, it would be useful to model the same system with two or more different techniques in order to investigate the respective strengths and weaknesses of each. Unfortunately, no such example exists in the literature, possibly because the act of formulating a single model represents a great investment of time and skill. Miller et al (1978) have commenced to resolve this issue by individual applications of SAINT and the optimal control model to a remotely piloted vehicle simulation.

The distinction between the two fundamental approaches to modelling is not the only barrier to comparison, however. Within any one approach, comparative evaluation may still be difficult due to the fact that the criteria of performance may be different. This issue is not particularly significant within top-down models because, as discussed by Rouse (1980), certain models tend to be more appropriate for certain types of systems. It would therefore be unlikely for a designer to be faced with a comparison of the server characteristics of a system (using queuing theory) and the system's control dynamics. However, within the bottom-up approach, this issue does become relevant.

For example, it is possible that a system may be rejected on the grounds that the memory workload of the operator is excessive (using HOS) whereas a simple analytic model may suggest that operator reliability is too low. This problem may be overcome somewhat if one specifies identical criteria for each model (a logical choice being 'time-on-task', due to the fact that all models may forecast this variable. However, a second problem may still arise, which is that the tasks which constitute each model may not be comparable. In particular, HOS relies on analysis of the system into 'basic procedures' rather than the discrete tasks of network models (Wherry, 1976). Even within network models, the variety of precedence relationships that may be employed in formulating a task sequence are a potential hindrance to model comparison.

The overall conclusion which may be drawn from this detailed technical discussion is that empirical methods of comparison may be extremely difficult to apply. Consequently, the favoured means of comparison has been logically-based, i.e., authors tend to compare models on the grounds of validity, generality, data requirements, etc. As models tend to answer different questions about the system, the best guideline that can be given about model selection is somewhat trivial, i.e., the model which is most appropriate for one's purposes should be chosen, provided that it can be applied in a tractable way.

4.7.2 Comparing systems for procurement

Whilst models may be used to verify that a system is in fact operable by humans, a second purpose of modelling may be to assist a decision about the purchase of competing systems. This issue is particularly relevant within the Australian procurement context, given our tendency to select systems from among overseas contractors. A related use of modelling may be to assess the worth of system re-design (when contemplating automation for example), and much of the literature deals with this topic.

The issues that emerged from the preceding discussion of model comparison apply equally well to the assessment of competing systems. In principle, the same type of model should be utilised for forecasting performance of each system. The criteria of performance should also be identical. If a bottom-up model is used then the structure of the task sequences may be different from system to system. Any decision concerning a comparison of systems based on data produced by a model will have to question in detail the validity of the task sequences used.

There are perhaps two approaches that could be used to make the comparison easier. The first of these would be to ensure that the bottom-up model includes system performance parameters. Thus, despite a difference in task networks between the two systems, global estimates of performance would be available and should be comparable.

A second method that could be applied is to make use of a scenario which standardised the task sequence as much as possible. The scenario could at least ensure that functions common to both systems were a requirement, despite task variation at a molecular level. (For a further discussion of the use of scenarios, see Section 7.2.3.)

4.7.3 Conclusions

While performance modelling has frequently been advocated as a means of systems evaluation that may be both timely and relatively inexpensive within the development cycle, it is certainly an underutilised technique. The use of prototypes and mock-ups are far more widespread. Undeniably, it is difficult to transfer system functioning into a relatively abstract performance model than into a more concrete prototype. This difficulty is compounded when dealing with computerised systems, such as C^2 systems. On the other hand, the benefits which may accrue from performance modelling seem too great to ignore. As a consequence, the pragmatic solution has tended to be the neglecting of model refinement in favour of obtaining approximate forecasts.

In Section 3.2, a number of pragmatic criteria for model evalution were proposed and Table 8 summarises our conclusions, based on inferences drawn from the literature.

A number of qualifications are required.   First, the evaluation neglects theoretical criteria such as validity, generality and parsimony in a direct sense.   Here we have taken a designer's viewpoint and presumed that validated examples of all these models would be available, although it is the case that some models are at a more advanced stage of refinement than others.

Secondly, the 'unknown' classification has been used quite liberally in Table 8, due to the fact that this report stems from a literature review.   In the case of SAINT, for example, it is known that the model incorporates parameters for individual differences and team behaviour, but no demonstration has yet been found of the application of those facilities.   Similarly, the relevance of top-down models at the early stages of design is unresolved.

Lastly, we have neglected the pragmatic concerns about the degree of effort or cost involved to apply a human performance model.   We also recognise that the 'transportability' of a model (Siegel and Wolf, 1981) is a significant issue, i.e., persons other than the model developers themselves should be able to use the model.   One reason for that neglect is because this report adopts a procurement viewpoint, in which case the onus to model is frequently with the contractor.   Additionally, data regarding the number of input parameters, programming expertise required, etc, are invariably scarce within the open literature.

Adopting the terms of our comparative discussion of human performance diagrams (see Section 2.11), we regard the relevance of models at early stages of design, and the relevance of models to computer system design, as the primary criteria.   From Table 8, it is apparent that no model satisfies both these criteria simultaneously.   Simple analytic models, such as those associated with reliability tree techniques, have the greatest application at preliminary stages of design, but fail to capture the nature of the human-computer interaction.   In fact, it is doubtful whether any of the models that we have reviewed adequately capture that interaction.   Both HOS and queuing theoretic models show potential in that respect, but empirical evidence is lacking at present.   The only model which has been specifically developed for computer system design, NETMAN, has been developed for design of a particular system and probably requires significant modification to become a general design tool.

Once again, it is not surprising that models applicable to the preliminary design phase are not also applicable to the development of computer systems, due to the difficulty of forecasting cognitive variables from an unembellished design.   The pragmatic solution would be in order, i.e., gross but timely models should be followed by more refined methods as development proceeds.   Secondly, information which may be gleaned from previous systems would be invaluable for promoting an appreciation of cognitive variables before a design project became advanced.   Computer systems are rarely developed in a vacuum and, in fact, often supercede manual systems, e.g., Pew et al (1978).   Analysis of the behaviour and requirements of operators in the current system may then form the basis for a systems model.   Lastly, the need for a precise model of human-

computer interaction may be reduced through the utilisation of reconfigurable software. In such cases, the need for accurate systems evaluation at an early phase may be somewhat offset by the capacity of the system to be re-designed during its operational life.

Early design	Relevant for application	Simulation human-computer interaction	Forecast of design solutions	Accommodates personnel requirements	team behaviour
a. Simple analytic models	H	L	L	L	L
b. Siegel & Wolf Naval models	M	L	M	M	M
c. NETMAN	?	H	M	M	M
d. SAINT	M	L	H	?	?
e. HOS	L	P	H	L	L
f. Optimal control model	?	L	H	L	L
g. Queuing models	?	P	H	P	?
h. Composite models	?	L	H	L	L

H = High utility
M = Medium utility
L = Low utility
P = Potentially useful
? = Unknown

TABLE 8 - Comparative summary of human performance models
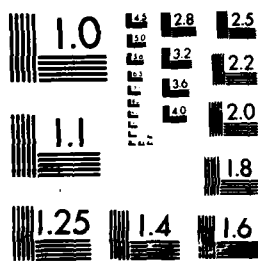
5. EXPERT JUDGEMENT IN SYSTEMS EVALUATION

5.1 Evaluation of Operational Systems

Both operational or prototype systems may be evaluated using either formal or informal techniques. Formally, systems testing consists of the collection of performance data under a representative sample of conditions (Meister and Rabideau, 1965). Such testing should decide the issue of whether system goals are being met, or are likely to be met in the case of a prototype system. For example, the effectiveness of a C^2 system may be determined by measuring such variables as mission time, accuracy, etc.

From a human engineering perspective, the major issue in systems evaluation is whether the demands of the system within various missions exceed operator capability (Meister, 1971a). If so, it may be said that a factor that contributes towards system ineffectiveness has been identified. Operator capability may be related, in a formal sense, to the ability to maintain performance within pre-defined limits of speed or accuracy. On the other hand, it is possible that performance may be within limits, but at the cost of relatively great operator effort. In such cases, operator opinion of the system (or prototype) may be sought as a significant means of evaluation. The use of such judgemental data constitutes a less formal type of evaluation.

Expert judgement can also be used as a valuable auxiliary form of evaluation. In the most extreme case where resources do not permit the application of other techniques, systems testing may be dispensed with altogether in favour of evaluation via an expert panel. However, this is unlikely to be the only procedure adopted in the design of complex systems for the armed forces, as prototype testing is usually a contractual requirement.

Operator and/or expert opinion of prototype systems has become a standard means of evaluation in at least two areas; namely aircraft handling quality and acceptability of decision-aiding systems. In the extreme case, the evaluative procedure consists of deriving a response concerning whether the system is acceptable or not. More frequently, operators are required to rate the system on pre-specified dimensions.

As regards aircraft, Knowles (1967) obtained rankings of six alternative flight control systems along five dimensions in order to assess their desirability. An interesting feature was that the opinions of both research pilots and human factors engineers were sought, and differences between the groups were noted. The well-known scale of Cooper and Harper (1969) applies a slightly different methodology in that the subject makes a series of yes/no decisions that culminate in a numerical rating of aircraft handling quality.

The acceptability of decision-aiding systems is of particular concern because of the variety of idiosyncratic decision styles that can exist. A further reason for concern is that the aids may prescribe a strategy which is in conflict with the sub-optimal biases which people employ, as described by Tversky and Kahneman (1974). Operator opinion has been used to endorse management systems in advanced aircraft (Steeb, Chu, Clark, Alperovitch and Freedy, 1979; Chu, Chen, Clark and Freedy, 1982), anti-submarine warfare (Leal, Chen, Gardiner and Freedy, 1978),

military tactical engagement (Kibler, Watson and Kelly, 1978) and for information flow in C^3 systems (Samet, Weltman and Davis, 1976; Samet and Davis, 1977). Operator opinion has cast doubt on the suitability of a system for remotely piloted vehicle control (Steeb, Chen and Freedy, 1977) and for option generation within C^2 (Tong, Arbel, Cioffi, Kelley, Payne and Tse, 1983).

5.1.1 Human Factors Checklists

Structured questionnaires and rating scales are also less formal techniques of systems evaluation when contrasted with those of diagramming and modelling. The utility of the former techniques, however, may be sufficiently high for them to influence design decisions. Human factors checklists, a long-established means of evaluation, may also be placed in this general category, due to the fact that they utilise subjective data in a 'paper and pencil' fashion. However, as noted by Siegel et al (1975), more sophisticated psychometric techniques may have a potentially greater contribution to make to systems design.

Checklists typically deal with fairly gross aspects of systems for which the evaluative criteria are well-documented. For example, it may be an obvious (but necessary) step to check whether an operator may reach all controls. The lists also tend to have a fairly strong physiological or anthropometric bias, which limits their use. It is entirely possible for a system to conform to such requirements but still be unacceptable. For these reasons the adequancy of human factors checklists such as those based on the U.S. Military Specifications (Mil. Specs) has been widely criticised, e.g., U.S. General Accounting Office (1981), although checklists would appear to have an effective role in screening near-operational systems.

5.2 Evaluation of Conceptual Systems

In a similar fashion to the preceding discussion, a distinction may be made between formal and informal means of evaluation at early stages of design. In the absence of at least some form of mock-up, actual performance data cannot be collected. At such stages of development, the most formal type of evaluation is through systems diagramming or modelling. This family of techniques allows projections to be made of system and, particularly, of human performance. They permit the comparative evaluation of alternative designs in order that "what if" questions may be answered. A secondary advantage is that they make the projected functioning of system (and operator tasks) explicit, which aids professional communication.

Within these specialised forecasting techniques, however, a certain degree of subjective opinion is usually present. In particular, many models require the assignment of values to input parameters that reflect human performance. Network models (see Section 4.2) especially rely on the synthesis of performance data for sub-tasks in order to yield the overall predictions. If the values of these parameters have not been established by previous research, then expert consensus is the most appropriate method. A second common use of opinion occurs within reliability modelling, in which the consequences of any one human error on total system performance may have to be estimated. Both Swain (1964) and Pickrel and McDonald (1964) have advocated the use of subjective data for this purpose.

In addition, the outputs of most models do not themselves provide a direct evaluation of the system, i.e., some form of interpretation is necessary. Common dependent variables of models are time-on-target and percentage of tasks (or missions) successfully completed. Using these data alone, alternative systems may be ranked according to their effectiveness. It then remains for the system developer to decide what constitutes acceptable levels of performance. In some instances, mission requirements may dictate the levels of performance required. In other instances, the values of these criteria may have to be derived by group consensus.

Aside from the subjectivity that is inherent in most specialised forecasting techniques, it is often considered legitimate for expert panels to evaluate conceptual systems directly. Once again, this judgement may be of a binary nature, i.e., whether the system is acceptable or not, but such judgements are more structured along a number of pre-determined dimensions. In particular, judges are frequently required to evaluate systems with respect to a number of pre-determined dimensions. An example comes from the work of Potempa, Lintz and Luckew (1975), in which judges evaluated the overall maintainability of certain systems by giving ratings of accessibility, complexity, etc. These dimensions of evaluation were established by a review of handbooks, operator interview, past research, and by prior 'expert' judgement. In other studies, the dimensions have been established by correlational analysis of questionnaire data (Topmiller, 1964; Knowles et al, 1969; Siegel, Fischl and MacPherson, 1975) or of performance data (Landis, Slivka and Jones, 1967). The techniques of multidimensional scaling and factor analysis have been most widely used for this purpose.

Having obtained ratings or rankings for several dimensions, weightings are then given to each dimension in order to arrive at an aggregate score for the system. These weights may once again be given by prior expert review, but have also been assigned with more methodological rigour. For example, Siegel, Miehle and Federman (1964) employed the mean weighting obtained from a judging panel. In other circumstances, objective evaluative data have existed that could be correlated with the subjective judgements. For example, whilst requiring maintainability evaluations from their panel, Potempa et al (1975) also had available maintenance time data for various subsystems. Under the assumption that the contribution of the maintenance factors was additive in nature, it was possible to construct a multiple regression equation that optimised the predictive power of the maintenance factors by differential weighting. A similar procedure was followed in the studies of Topmiller (1964) and Landis et al (1967). The methodology of Siegel et al (1964) was unusual in that a multiplicative relationship between dimensions was assumed.

5.2.1 Visual display design

More specifically, there appear to be three areas within the domain of the human engineering of conceptual systems that have utilised subjective data extensively. The first area is that of visual display design where such evaluation is based on the belief that 'critical parameters' of display effectiveness exist (Landis et al, 1967). If these parameters can be identified and judged consistently, then an evaluative technique exists that can be used instead of constructing models and mock-ups. Attempts to devise such a technique have included

the Display Evaluative Index (DEI) of Siegel et al (1964); the Analytic Profile System (APS) of Siegel et al (1975) and the Decision Quality Metric (DQM) of Landis et al (1967). Two of the metrics, namely the DEI and the DQM, in fact require input data that are of variable 'subjectivity'. For example, one assumes that a judgement of display size (in the DQM) is more amenable to external scrutiny than a judgement of mismatch between information required and displayed (in the DEI).

Validation of these techniques of display evaluation has revolved around two areas: demonstrations that the techniques have desirable psychometric properties, and efforts to show that the value of a display index can be a good predictor of human performance when humans must process information from that display. In effect, these techniques of display evaluation utilise the judge as a measuring tool. It is therefore a desirable psychometric property that the resulting judgements of any individual should be consistent, i.e., reliable across time, and reasonably sensitive to differences in displays. For example, the fact that the overall APS score is based on forced-choice comparisons is said to be important in this regard.

The predictive validity of both the DEI (Siegel & Federman, 1967) and the APS (Siegel et al, 1975) have been evaluated. The general methodology used in such evaluation has been to construct alternative versions of a large-scale C^2 display and then to derive performance scores for processing of the displayed information. For example, in the study of Siegel et al (1975), subjects were required to make tactical decisions that could be graded as 'correct' or 'incorrect'. The total APS index was shown to be a satisfactory predictor of those scores, even when employed by naive specialists.

5.2.2 Maintainability

System maintainability is another domain in which expert judgement has figured prominently in the literature. Both Topmiller (1964) and Potempa et al (1975) have constructed multiple regression equations for prediction of U.S. Air Force system maintainability from a human factors perspective. These equations are analogous to a simple additive factor maintainability model that utilises subjective input data. Whilst these predictive techniques have been shown to be reasonably valid for the particular sub-systems on which they were devised, their generality has not been established.

5.2.3 Personnel forecasts

Projections of manpower and training level have also been subject to panel review. Both Potempa et al (1975) and Rossmeissl and Dohme (1983) have made preliminary attempts at using rating scales to determine aptitude requirements of Air Force and Military systems, respectively. Problems with making manpower estimates at early stages of design have been described by Gael (1964).

One significant problem is that the configuration of the system is liable to change frequently, particularly at conceptual stages of design. Additionally, the manpower estimates may be used actually to determine the final manpower structure rather than being used as a predictor or check of a system's configuration.

Whilst the field of personnel forecasting is one which is logically suited to expert judgement during systems design, relatively few such studies exist in the literature. It is probable that further research is needed before a coherent set of valid techniques may be said to exist. The impact of a system upon personnel resources is a perennial design issue, and further discussion may be found in Section 6.15.

5.3 Summary of Expert Judgement Techniques

While some system developers may object to the use of subjective data as being unreliable, Siegel et al (1964), Landis et al (1967), Knowles et al (1969) and Siegel et al (1975) have demonstrated how statistical techniques may be used to assess such data. As a minimum, experts should show consistency in their judgements across time and should be sensitive to differences in the systems (or the aspects of the systems) that they are required to judge. Additionally, the finding that reasonably large inter-judge differences are often found poses a significant problem (Knowles et al, 1969). As discussed, the use of pre-defined dimensions and weighting schemes assists in the standardisation of expert opinion. However, there is still no guarantee that the ratings or rankings obtained from judges will be uniform. One solution, as suggested by Knowles et al (1969), is that empirical data may be used to identify the most valid judges. In other circumstances, the composite judge (representing the mean group rating) may be the most valid.

The process of arriving at group consensus can also be structured. De Greene (1970) has recommended the Delphi technique for shaping expert opinion during systems design. In its most basic form, this technique consists of a successive modification of individual opinions by exposure to the opinions and presumptions of the group as a whole. The facility for revision of opinion under the influence of group feedback, possibly in an anonymous fashion, is the principle on which the technique is based. The Delphi procedure has been developed largely as a means of enhancing group forecasts of a political or economic nature, but in principle should be useful during systems design. Siegel and Wolf (1981), in particular, have speculated that the technique may be useful for parameter estimation when developing human performance models.

In conclusion, a number of techniques exist for structuring the subjective evaluation of conceptual systems. The most useful of these have been shown to be psychometrically reliable and to have predictive validity. A principal value of these techniques may be that they require a multidimensional analysis of the system. Therefore, not only may judgements of the overall acceptability of certain systems be assisted, but design improvements for various system aspects may also be stimulated. The human factors checklist as an evaluation tool needs to be applied with caution.

6. HUMAN FACTORS ISSUES IN SYSTEM DEVELOPMENT

6.1 Introduction

One of the goals of this report is to discuss some design issues that occur frequently in the system development cycle. Particular attention has been reserved for issues which arise in the development of interactive computer systems, as those issues are most relevant for the design of RAN combat data systems. A list of the issues which this report addresses was given previously in Figure 1.

The purpose of this discussion is three-fold. First, the definition of certain issues has the desirable effect of heightening the awareness of designers who may have otherwise overlooked such factors. Simplistic though this 'consciousness raising' approach may be, it can be important because design issues are rarely specified in the development contract (although they could be). Rather, the goals of the system tend to be given in very global terms, together with some cost constraints. It is then left to the contractor to design a system which will meet those goals in any manner that is thought fit. Some of the design issues to be considered here have been discussed previously in standard human engineering textbooks, but there is no guarantee that those textbooks will be consulted by system designers. Some issues are deemed to be of sufficient importance to emphasize here.

A second purpose of this section on design issues is to provide a review of how a number of applied human factors methods have been used to resolve these issues in the past. As discussed in the introductory stages of this report, recognition that certain design issues exist tends to lead to the recognition of a design 'problem'. For example, recognition of vocal vs non-vocal modes of communication as an issue leads to a consideration of the merits of different systems which employ those modes. A number of techniques may be used to evaluate alternative system concepts from a human factors perspective in order to choose the most effective, and thus resolve the design problem.

There is a distinction between those techniques which are applicable at preliminary stages of design and those which may only be applied when the design is more detailed. At preliminary stages of design, an understanding may be gained of human performance within the system through drawings and models, in a way that will aid evaluation. At detailed stages of design, the system is more likely to be evaluated through a prototype test. Although the philosophy underlying this report has been to review design methods at relatively early phases of development, those issues which are potentially open to resolution through the use of mock-ups, prototypes, etc., have been included for discussion.

The last purpose of this section concerns issues that are sufficiently concrete and well-researched for standard design guidelines to exist. Where possible, these guidelines have been paraphrased for the benefit of system developers and/or procurers. However, it should be noted that the major objective of this report has not been to provide data about specific design issues. Rather, the report primarily concerns the provision of guidance of a more conceptual nature. Accordingly, the search for design 'rules' has by no means been exhaustive, and the reader is invited to consult the references on particular issues for further information.

A brief explanation regarding the choice of design issues in this report also seems to be in order. Most issues were identified during a review of the literature pertaining to human factors evaluation at early stages of design. Some issues have been included not because they are particularly tractable using those methods of evaluation, but because they have been recognised as important within military systems development. Once again, it should be stated that a complete literature search and review was beyond the resources available in this project. The review requires augmentation before a coherent set of design recommendations may be assembled. For an authoritative example of how empirical studies may be related to design recommendations, Meister (1976) is a useful reference.

Viewing the list of issues given in Figure 1, it is apparent that not all are of the same type. That is, while all qualify as design 'issues', there is a range in the degree of abstractness between them. Human-machine allocation of function, as previously discussed, is one of the initial human factors related issues that should be dealt with early in the system development cycle, and consequently has a pervasive influence on the subsequent development. The merits of graphical and textual modes of information presentation, on the other hand, may be regarded as a relatively concrete software decision that occurs at an advanced stage of design. Other issues, such as the effects of operator strategies or stereotyped behaviour upon system performance, have only an indirect relationship with the system design process. Conventionally, these have been regarded as organisational or training isues.

6.2 Human-Machine Allocation of Function

In hardware systems design, the major concern at the predesign phase of development is the allocation of function between humans and machines. That is, as it is presumed that system functions are independent of each other to some extent, a reasonable issue concerns which of those functions should best be automated. The basis of this decision may be a formal comparison of the abilities of people and machines (such as the so-called Fitts list). However, this process has been criticised on the grounds that it neglects the fact that people and machines are complementary (Jordan, 1963). Secondly, the comparison may be misleading, for cost constraints or technical feasibility may dictate that the allocation of function is less than ideal (Chapanis, 1965).

All decisions regarding allocation of function imply that some form of systems evaluation has been carried out with human engineering criteria in mind. That is, if alternative system concepts are entertained, making the choice of one implies that its performance has been hypothesised to be superior (after taking into account the costs involved). Even if alternative concepts are not considered, it may be said that performance has been hypothesised to be satisfactory in relation to system goals. Thus, many human factors inputs are preventive in nature; the value of these early performance forecasts does not become apparent unless human factors are ignored, leading subsequently to design problems.

The means of making predictions of what problems may arise, and how they can be avoided are often obscure, particularly as human performance requirements tend not to be articulated in detail at the

planning stage. That is, it is rare for human performance to be predicted and then compared with a criterion value. Rather, a gross estimate of human performance is made and then the effects of this performance upon system effectiveness are estimated. If human performance is predicted to be unsatisfactory, then it may be said that the design has exceeded operator capability in some respect. The most commonly employed parameters of human performance are speed and accuracy, and thus a high priority is placed on the identification of situations of excess 'workload' or 'stress'. In other circumstances, it should be noted that such factors as the safety of the operator may be of greater concern than operator capacity, although both factors have much in common.

In practice, the textbook case of how allocation of function occurs is probably somewhat fictitious. Designers tend to make an initial allocation based on past experience, hardware cost/availability, and possibly some generalisation such as "people have the advantage of flexibility but are not as reliable as machines". This initial allocation is then retained unless later analysis proves the decision to be unsatisfactory. In particular, if operator capability is discovered to be insufficient at a later (less abstract) stage of design, a re-allocation of function may be necessary to alleviate this problem.

Functional allocation decisions tend to have a profound influence on system design. From a human viewpoint, once a certain function has been assigned to a machine, the flexibility of behaviour of the humans within the system has been reduced. Correspondingly, the options for solution of a given design problem become constrained once a functional allocation has been made. For example, identification of a 'sensing' function with 'radar' automatically defines the roles which some operators will play within the system, and suggests that all further design decisions will be concerned with embellishment of that configuration. In the development of systems that are variations on a familiar design, many allocation decisions may be pre-determined. However, in the design of relatively unique systems, such as many computer-based systems, functional allocation assumes special importance (Kidd and Van Cott, 1972).

In a technical sense, many of the detailed interface issues which occur during system development may be interpreted as functional allocation decisions. That is because what interface alternatives are possible depend directly on the degree of automation available. For example, a choice between an analogue and a digital read-out may appear to be straightforward until it is realised that analogue displays give additional information about the rate of the quantity which is being measured. This could imply that the human has been assigned an extra monitoring function.

It is possible to distinguish three broad philosophies regarding allocation of function (Nickerson, Meyer, Miller & Pew, 1981). The first, which is possibly the most common, specifies that all functions should be automated whenever it is possible to do so at a satisfactory cost. The support for this philosophy is often based on arguments that automation reduces the potential for human error, and may also reduce personnel costs. However, automation has often resulted in unforeseen maintenance costs (Smith, 1980). Furthermore, the advantage of reduction in human error may be achieved at the expense of diminished system flexibility, due to the fact that the unique adaptive talents of the human have not been utilised.

In the extreme, the functional allocation approach based on the adaptive capacities of the human specifies that all functions should be manual unless there are compelling reasons to the contrary. This view has been articulated, with some qualifications, by Meister (1971a). (Examples of 'compelling' reasons would be that long-distance surveillance obviously requires technical implementation, or that some functions may be so critical that no human error may be tolerated).

The third philosophy represent the compromise position, i.e., the majority of functions should be semi-automatic. This view is most congruent with the interactive system design philosophy that requires that human and computer should co-operate on most tasks. For example, a common adage is that the computer should perform routine clerical and arithmetical work, whilst the human performs most decisions within a problem-solving task (Licklider, 1960). Such a view rests implicitly on a comparative analysis of the abilities of people and machines which, as has been discussed, may obscure the complementary nature of the human-machine relationship. On the other hand, the objection is not so much against the comparison itself, but against the ends to which that comparison may be directed.

In a more formal sense, functions may be allocated by forecasting (in a quantitative fashion) the human and/or system performance that is likely to result from a given allocation, and then deciding whether that allocation is satisfactory. At early stages of design, the most useful techniques for aiding performance forecasts are diagrams and models, and the following review will concentrate on design studies which have employed such techniques. In fact, an argument could be made that if functional allocation is verified solely through a hardware mock-up or prototype, that evaluation has occurred too late in the development cycle for re-design to be convenient or inexpensive.

Both person-machine and person-person allocation of function appear to be design issues that are well-suited to resolution by conceptual means of systems evaluation such as diagrammaing and modelling. For example, Lindquist et al (1971a), during the design of a tri-service search-and-rescue helicopter, derived eight common functions, viz: flight control, navigation, communications, surveillance, systems monitoring, environmental sensing, search and rescue. There was a logical order and a predictable duration of these functions. From the functional diagram, it was possible to specify the tasks of the crew and represent these along a time-line. Various allocations were tested in order that crew workload should not be excessive. That is, if the predicted duration of certain tasks was greater than the time which was available through mission constraints, a re-allocation of tasks amongst the crew was made. Some procedural modifications were also used as design solutions.

Siegel, Leahy and Lamb (1976b) employed their "4-20 man" network model in the evaluation of a naval sonar system. Briefly, the system was relatively close to operation (at the detailed design stage) and the modelling was used to suggest design modfications. In particular, the supervisor of the sonar operators was shown to perform in a fashion that would have been reasonably expected to degrade system effectiveness. The solution was that an automated decision aid of an unspecified nature might alleviate the workload of the supervisor. Reallocation of function amongst the operators was also suggested as a means of reducing the load.

A conceptually distinct model of Siegel et al (1981), called NETMAN, has also addressed the allocation of function issue. This model has been developed to aid the on-going design of an army field exercise management system. The approach is unusual in that communication aspects of the system receive emphasis. The model was originally developed on the TWSEAS system, but has subsequently recommended that a modified system, EMARS, was more effective. The new system was more highly automated, though few details are available.

As regards information processing models, HOS provides an example of the use of performance modelling to evaluate the effect of level of automation on system performance. There were two parts to the study by Lane et al 1980 which was concerned with the crew-station onboard a U.S. Navy anti-submarine aircraft. First, the modelling was used to demonstrate that the presence of a Forward Looking Infra Red (FLIR) system was causing the crew excess workload, although this problem was also recognised from operational conditions. Next, the model facilitated suggestion of a solution involving automation of some of the navigation functions as compensation for performance demands of the FLIR system.

The optimal control model has also been used to assess the merits of automation to the system. Hess (1981) demonstrated that an automated flight director system in a helicopter could be expected to improve system performance.

Surveying the range of studies that have attempted to resolve functional allocation issues via modelling, it is conspicuous that a technical solution has been the most frequent recommendation. That is, it has usually been hypothesised that system performance would be increased through the automation of one or more manual functions. Most modellers have subscribed to such a technical philosophy of functional allocation with little regard for the potential disadvantages. The reasons for this trend are probably complex, but speculation is possible. First, most human performance modelling, to be productive, is carried out for functions which are crucial or time-stressed, e.g., Siegel et al (1977). Priority is therefore placed on the reduction of human error through the overcoming of human limitations. Secondly, most modelling situations are contrived so that human behaviour is restricted and paced by the demands of the system, which means that the uniquely human talents of flexiblity may be neglected to some extent.

However, it cannot be denied that an underlying philosophy of conventional human factors engineering is design for the majority of the user polulation, so that design problems tend not to be resolved solely through the institution of training programs or other specialised procedures. This philosophy has been reflected in studies by Akst (1982) and the NTDS Functional Allocation Study Group (1981), in which automation was explicitly anticipated as the prime means of improving the present manual C^2 system (within the U.S. Marine Corps and Navy, respectively). In that context, the greatest concern was for the cost of various levels of automation.

Some exceptions do exist. For example, Lane et al (1980) demonstrated through their model that the addition of an FLIR system actually had deleterious effects upon anti-submarine warfare performance. Siegel et al (1976a) did recommend a training programme as

a possible means of improving sonar operator performance, whilst both Lindquist et al (1971a) and Siegel et al (1976b) opted for a reallocation of tasks amongst the crew members as a means of equalising the workload. Montgomery, Thompson and Katter (1980), in analysing a U.S. Army intelligence system, stated that one of the goals of their report was to reorganize procedures within the system in order to enhance performance. Jorgensen and Strub (1979) investigated both team structure and procedural factors through a prototype of the U.S. Army AN/TSQ-73 air defence system, as well as making recommendations for automation.

6.3 Decision-Aiding Systems

6.3.1 Introduction and definition

In a broad sense, any device that assists an operator in the making of a decision may be regarded as a decision-aid. Therefore it may be legitimate to characterize reference manuals or other job aids in this way. However, for the purpose of this report decision-aiding systems will be considered to be mediated by a computer and some form of interactive interface with the decision maker.

While some preliminary definitions are being made, it could also be said that virtually any computerized system functions as a decision-aid. That is because a major role of computer systems is the storage and transmission of information in order to assist the user with some task. This emphasis on decision-aiding is particularly relevant if the task is managerial in nature, i.e., if the task involves some form of problem-solving. Such tasks are commonly associated with commercial business strategy but, in a military sense, the tactical C^2 situation is equally appropriate.

Historically, computerized management systems have evolved through distinct phases (Brookes, Grouse, Jeffrey & Lawrence, 1982; Tucker, 1980). The initial phase was associated with the advent of 'management information systems', in which information flow and transaction processing became automated. More recently, 'decision support systems' have emerged in which the actual managerial process is becoming subject to computerization, thus satisfying the conventional definition of a decision-aiding system more exactly.

The goal of decision-aiding, in the words of Crolotte and Saleh (1980), is "to allocate information processing and decision functions between man and machine in a way which optimizes the use of their respective strengths and compensates for their respective weaknesses" (p. 1069). Special attention is usually reserved for the weaknesses of the human. In particular, aids should circumvent the human's inherent cognitive limitations, such as those pertaining to memory, and should overcome the human's deficiencies, such as decision-making biases. Transitional management information systems have partially resolved the former problem, while the newer decision support systems are starting to address the latter problem as well.

The issue of whether or not to implement a decision-aiding system may therefore be regarded as a special example of functional allocation. Decision-aiding systems invariably require the (partial) mechanization of what were previously 'manual' decision functions, and thus the perennial issue of the relative merits of human and machine

applies. The issue is less straightforward than those previously encountered (if any functional-allocation decision may be termed straightforward) because many of the processes involved in managerial decision tasks are covert and difficult to analyze.

6.3.2 Decision tasks and taxonomies

The difficulty of analyzing decision processes is one factor which, historically, has inhibited the design of decision-aiding systems. However some headway has been made in decomposing decision-making tasks into sub-tasks so that they may be subsequently classified. Techniques such as the analysis of verbal protocols (Triggs, 1973) have proved to be especially useful in this regard. It has thus become possible to construct a decision-task taxonomy.

A variety of decison-task taxonomies exist, according to the purposes of the individual researchers. A typical example is shown in Table 28, taken from Crolotte and Saleh (1980). That taxonomy makes a distinction between the attributes of a decision task (i.e., the conditions under which the task is performed) and the functional requirements of a decision task (i.e., the formal steps involved). For the purposes of this report, the latter dimension alone will be discussed. A more comprehensive taxonomy along that dimension is shown in Table 29, taken from Nickerson and Feehrer (1975). However, even that taxonomy may be criticized on the grounds that it neglects a phase of option generation, which Wohl (1981) believes is crucial within tactical C^2. Wohl's (1981) stimulus-hypothesis-options-response (SHOR) taxonomy of U.S. Air Force decision-making is shown in Table 30.

Different types of decision-making tasks are believed to be distinguished by the fact that they emphasise different steps in the decision-making process. In a military context, command and control (as a general process) involves all of those stages. However, sub-functions of C^2 may be further distinguished. For example, it is the case in intelligence analysis that a high premium is often placed on developing hypotheses about the state of the world, due to low quality of input data, whereas the available actions may be clearly prescribed (Wohl, 1981). In other circumstances, the input data may be relatively unambiguous and the major decison is one of response selection, e.g., allocation of weapons to a set of known targets.

Attributes	(Single attribute/multi attribute
	(Individual/group
	(Static/dynamic
	(One shot/repetitive
	(Certainty/risk (uncertainty)
	(Abstract (general)/concrete task specific
	(Well defined/ambiguous
	(Time critical/time relaxed
	(Small probability high loss/normal range

Decision Task

Functional Requirements	(Problem recognition
	(Alternative development
	(Information acquisition and evaluation
	(Alternative evaluation/selection
	(Feedback monitoring

TABLE 28 - Decision task taxonomy (Crolotte & Saleh, 1980)

a.   Information Gathering

b.   Data Evaluation

c.   Problem Structuring

d.   Hypothesis Generation

e.   Hypothesis Evaluation

f.   Preference Specification

g.   Action Selection

h.   Decision Evaluation

Table 29 - Classification of decision situations or tasks (Nickerson & Feehrer, 1975)

In a corresponding fashion to the variety of decision-task types a range of decision-making aids also exist. Some aiding systems in fact constitute a number of separate decision aids which may be applied to different tasks. More commonly, decision aids tend to be specific to one or more of the stages of decision-making that have just been described. For example, the well-known Bayesian algorithm is best suited for assisting the probabilistic evalution of hypotheses, whilst multi-attribute utility theory provides a rational method of preference specification.

A recent taxonomy of decision aids, together with the decision-making stages that they assist, may be found in Crolotte and Saleh (1980). Given the orientational nature of this discussion, no attempt will be made to analyze the details of that taxonomy. However, at a more general level, Zachard (1980) believes that all decision aids fulfill one of seven functions, namely, outcome calculation, value specification, data control, data analysis, data entry or display, and human judgement amplification or refinement.

As mentioned previously, decision-aiding issues may be interpreted as special cases of functional allocation. Given that decison-aiding taxonomies exist, the allocation of function problem reduces somewhat to one of analyzing the decision-making task, then matching the task to its appropriate aid. Unfortunately, the efficiency of many aids has not been established; for example, even the well-researched Bayesian approach is subject to many limitations (Bowen, Feehrer, Nickerson, Spooner & Triggs, 1970). In the case of C^2 generally, there are compelling reasons for believing that a number of aids might be useful, so that priority would be given to decison-tasks (and their respective aids) when considering a system re-design, e.g., Zachary (1980). The attention given to the development of such aids depends somewhat on the degree to which automation is deemed desirable by the system developers.

6.3.3 Artificial intelligence

The definition of an artificial intelligence (AI) system is somewhat obscure, due to the variety of interpretations which different researchers have used. Broadly, AI techniques may be regarded as constituting a sub-set of decision-aiding techniques. However the distinction between AI and decision-aiding is yet to be articulated clearly. In fact, referring to Zachary's (1980) functional categorisation of decison aids (i.e., as those which either calculate outcomes, specify values, control data, analyse data, assist data entry/display, or amplify/refine human judgement), it is possible to find examples of each category in the AI literature.

For the purpose of this discussion, the framework adopted by Phelps, Johnson and Halpin (1979) is most suitable. These authors have made a distinction between information aids and integration aids within decision-aiding systems. Briefly, information aids are used for automatic data selection, or for the performance of calculations, whilst integration aids further transform data by organising and structuring information, by weighting information or by overcoming the biases and limitations of the human. It is the latter roles that probably have the greater correspondence to the popular notion of AI.

Generic Elements	Functions Required	Information Processed
	Gather/Detect	
Stimulus	Filter/Correlate	Capabilities Doctrine: position, velocity,
(Data)	Aggregate/Display	type, mass, momentum, inertia, relevance and
S	Store/Recall	trustworthiness of data
Hypothesis	Create	
(perception	Evaluate	
alternatives)	Select	
H		
Option	Create	
(response	Evaluate	
alternatives)	Select	
O		
	Plan	The air tasking order:
Response (action)	Organize	Who What When Where How How Much The near-real-time modification/update

Table 30 – Anatomy of tactical decision process – the SHOR model (Wohl, 1981)

Bechtel (1981) provides a good overview and critique of several AI systems relevant to C^2. Within a tactical C^2 context, two types of AI techniques have received relatively widespread attention (Vittal, Selfridge & Bobrow, 1981). These are knowledge representation systems and inference systems. Briefly, knowledge representation systems provide a framework for organising and storing information, according to some pre-defined plan. Such information may include representation of spatial data, and may be displayed graphically. Thus, these systems may permit a fast-time simulation of a tactical situation in order that projections about the future may be made. One function of knowledge representation systems is to prevent information overload during a crisis situation by representing information clearly and compactly. Inference systems, on the other hand, actually contain stored rules or algorithms that have

frequently been derived from military experts. Given suitable data, these systems may deduce conclusions about the state of a tactical situation, and may even recommend actions.

For knowledge-based systems, Pease (1978) has made some preliminary recommendations relevant to C2:

(i) Generally, such systems should be both flexible and adaptable;

(ii) The user (i.e., the commander) should be possessed of relatively great control of the system whilst requiring little knowledge of the technical details of system operation;

(iii) Data should be easily accessible;

(iv) A facility should exist for definition of potentially critical situations, using a natural language; and

(v) The scope of the system should be modifiable.

From these recommendations, three design principles have emerged, viz:

(i) Separation, i.e., the different features of the system should be easily distinguishable. Within computer science, modularization of software achieves this goal;

(ii) Similarity, i.e., the system should appear to operate in the same way that a comparable all-human system would operate; and

(iii) Negotiation, i.e., there should be a means of co-ordinating the operation of various system modules.

6.3.4 Methods of evaluating decision-aiding systems

Decision-aiding systems are relatively undeveloped at present and there are consequently few methods of evaluation discussed in the above literature. In fact, some reviews have concentrated on a logical comparison of decision-aiding systems when empirical data have been absent (Bechtel, 1981; Siegel, Madden & Pfeiffer, 1980). In the majority of studies, however, decisions aids have been tested through operator interaction with a prototype.

A small number of evaluative studies using diagramming/modelling techniques do exist. For example, Siegel et al (1976 a and b) used their network model to recommend that decision-aiding might improve the performance of operators within a U.S. Navy sonar system (at the detailed stage of design). Tainsh (1982) has also speculated that his job process charts, that diagram human-computer interactions, may be useful for identifying the need for decision-aids. A similar claim has been made for the HOS model (Lane, Strieb & Leyland, 1980).

One criticism which may be made of these claims is that so far neither diagramming nor modelling has been used to make specific recommendations for a decision aid. That is, while decision-aiding may have been recommended as a means of alleviating operator workload, the form which that aid should take has usually been neglected. This limitation probably results from the relative crudeness of most efforts to model cognitive processes. A possible exception are IDEF diagrams (Wohl et al, 1983), which are claimed to permit the decomposition of individual

decisions, although more evidence is required before a firm conclusion can be reached. A related criticism of these studies is that they have often not revealed anything about the need for decision-aiding beyond the fact that a particular task is error-prone. In such instances, the value of the research may lie more in the background task analysis than in the modelling diagramming technique per se.

As an addendum, the relationship between control theoretic models and decision-aiding deserves attention. A fundamental postulate of control theory is that typically greater feedback enhances performance, subject to the caveat that multiple sources of feedback can cause decrements due to division of attention. Many studies using the optimal control model in particular have been devoted to assessing the impact of greater feedback (commonly through augmented displays) upon performance. To the extent that augmented displays (i.e., quickened, predictor, or higher-order displays) may be regarded as decision aids, control theoretic modelling is relevant to system evaluation. Studies by Connelly (1977), Hess (1981), Johannsen and Govindaraj (1980) and Kraiss (1981) have been discussed elsewhere in the modelling section of this report and provide examples of decision-aiding evaluation within military contexts.

6.4 Hypothesis and Option Generation

Hypothesis and option generation are the names given to two stages within the decision-making process. The SHOR taxonomy (Wohl, 1981) contains the most explicit military representation of these two stages (see Figure 30) . Hypothesis generation refers to the ability of the command, acting upon incoming data, to formulate ideas about the possible state of the world, e.g., "what could the enemy be doing?". Option generation refers to the ability to formulate alternative courses of action, e.g., "what can be done about this tactical situation"?

Hypothesis generation is invariably followed by a stage of hypothesis evaluation in all decision taxonomies, i.e., the probability of the various hypotheses must be ascertained. Similarly, option generation is followed by a stage of preference specification, in which the most desirable or effective option is selected. Laboratory research on decision-making has tended to be based on the assumption that hypotheses exist a priori and their evaluation has been investigated. Alternatively, the researcher has specified the available options and has then studied their selection. The result is that hypothesis evaluation and preference specification are relatively well-researched to the point that decision-aiding algorithms are available (in particular, those based on Bayes theorem and multi-attribute utility theory, respectively); whereas the initiating decision stages for both of those processes have been neglected.

Hypothesis and option generation have some degree of commonality in that both are essentially creative processes, in the sense that they require concept formation. Such a process has been difficult to investigate in basic research, and has prevented the development of suitable algorithms or heuristics. However, the importance of hypothesis and option generation to tactical C^2 has been repeatedly stressed, e.g., Nickerson and Feehrer (1975), Ramsey and Atwood (1979), Wohl (1981). For that reason, these crucial but least understood phases of the decision-making process have qualified as an issue that is essentially distinct from the decision-aiding considered in this report.

Unfortunately, few data exist with which to make design recommendations. Various studies of creativity and concept attainment have been performed, but these have been fragmentary and have lacked direct military application. In fact, only two relevant studies have been found during the limited review undertaken within this report, and both emphasised option generation. Both utilised a prototype as the method of investigation.

Gagliardi, Hussey, Kaplan and Matteis (1965) were concerned with the ability of operators to allocate missile-firing submarines to a set of targets. This allocation could be performed in a number of ways, so that an optimum response strategy had to be devised. Prior research had shown that subjects were inefficient in their search for solutions, so some form of decision-aiding was suggested. The task was well-structured enough for complete automation (through an algorithm) to be possible; however, it was found that the most effective system was semi-automated, i.e., a computer generated 'key elements' of the solution whilst the operator assembled those elements into a deployment. One criticism which may be made of this study is that the conditions were idealised (to use the terminology of the authors), i.e., the fact that the construction of an algorithm was possible suggests that there was little scope for the creative component of the option generation process.

Tong et al (1983), on the other hand, rejected the concept that option generation may be automated (at least in the real-world) and instead set out to create an 'option prompting environment'. The environment consisted of a complex of option prompting 'tools', derived from work in the fields of lateral thinking, brainstorming, etc. Basically, the purpose of the tools was to cause subjects to order their priorities and challenge their assumptions. Subjects were U.S. Naval postgraduate students who were required to respond to a tactical problem-solving scenario. The options which they generated were then evaluated according to a number of criteria such as breadth, novelty, feasibility, etc.. Qualified support was found for the use of an option prompting environment.

6.5 Decision Making v. Sterotyped Behaviour

As implied by the discussion of hypothesis and option generation (in Section 6.4), the various stages which are involved in making a decision possess quite different characteristics. In particular, some stages are more amenable to decision-aiding algorithms than others, due to the fact that sufficient theory exists for prescriptive recommendations to be made. The stages for which theory has aided prescription most prominently are hypothesis evaluation and preference specification, although even those prescriptions require well-structured conditions which may seldom be encountered in tactical C^2.

A second, more general, means of aiding decisions may be through the use of standardized procedures. In a tactical context, it may be possible to provide a course of action within any situation, based upon past experience. The decision-making problem therefore reduces to one of identifying the characteristics of a particular tactical situation and then matching the situation to the appropriate reaction. In fact, there is good evidence that such heuristics are often used, based on the 'commander's catechism' (Wohl, 1981).

To the extent that different problem-solving situations require different degrees of initiative and originality, a distinction may be made between decision making and stereotyped behaviour. The latter behaviour refers to those decisions which are primarily a result of previously-tested solutions. The greatest opportunity for stereotyped behaviour occurs through the use of standard procedures, although it is possible that some decision-aiding systems (including, particularly, artificial intelligence systems) may eventually automate much of the decision-making process. This latter possibility is small, however, for two reasons. First, such decision aids require well-structured situations to be effective, a requirement which is rarely satisfied in real-life. Secondly, it is doubtful if some of the more creative stages of the decision-making process (such as problem structuring, hypothesis generation and option generation) may ever be automated completely. At present, the most feasible human-computer relationship still appears to be that of Lickider (1960), i.e., people set the goals whilst the computer performs data aggregation.

Stereotyped decision-making behaviour has both beneficial and negative aspects. Military training may emphasise stereotyped responses through the study of previous tactical situations. Such behaviour has the effect of making manageable an otherwise confusing situation, and reduces the cognitive 'load' required to make a decision. However, for this procedure to be effective, a current tactical situation must have strong similarities with some previous situation, at least in a generic sense. If not, the command may lack the flexibility to make the appropriate response. Wohl (1981) has also commented that the pace and uncertainty of modern warfare places a high premium upon the more creative aspects of decision-making.

Given that standardized procedures may be effective in some cases, there may also be a problem with storing and retrieving that information. For example, Rouse and Rouse (1981) have emphasised both the time-stress involved in looking up emergency procedures for aircraft, and the bulk and inflexibility of that information. In addition, placing that information in the computer resolved some of those problems but created some new ones, such as the inability of novice operators to manipulate the keyboard.

6.6 Manual Back-Up

The issue of system operation during conditions of failure is particularly important within C^2, as there may be little opportunity for maintenance work during a battle. From a design or procurement viewpoint, therefore, the 'survivability' of systems is likely to be an important consideration (Goodbody and Monteleon, 1976).

Unfortunately, few design recommendations exist in order to ensure system survivability. General statements that the system should degrade gracefully (Vaughan, Whittenburg & Gillette, 1966) have the appropriate spirit, but are of little specific value. Possibly the strongest observation that has emerged is the advantage of distributed systems during times of crisis (Carley, 1967). That is, whereas a centralized system may suffer a decisive assault from which it may not recover, distributed systems may continue to function following damage to any one component. In this context, the term 'distributed system' may

refer both to (computer) systems in which the architecture is de-centralized, and to management practices in which the leadership function is de-centralized.

From a human factors perspective, another means of enhancing system survivability is to ensure that automated functions may be performed manually in the event of failure. That is, the allocation of function should be flexible as, incidentally, has been recommended for coping with variable workloads (Rouse, 1977). This issue is particularly relevant to computerized systems because, as the operator may function more as a supervisor and less as a direct controller (Nickerson et al, 1981), the ability to provide manual or backup support during times of failure is emphasized.

Two prototype design studies have been found during this contract which bear on the issue of manual back up. Jorgensen and Strub (1979) investigated the threat evaluation and weapons assignment (TEWA) function in the U.S. Army's AN/TSQ-73 air defence system. During simulated conditions of heavy load, i.e., with greater numbers of approaching aircraft, it was recommended that a fully automated system was necessary; however, manual operation was shown to be effective under moderate loads. Kriefeldt (1980) investigated distributed management for air traffic control. Background research had indicated that pilots preferred to have information regarding the trajectory of other aircraft available in order that they could initiate their own flight-paths. Such a system was proposed to replace the current system in which most traffic control was the task of a centralized authority. An interesting aspect of the study was that failure conditions were simulated, i.e., some pilots were permitted distributed management whilst others had to revert to the centralized system. The distributed system, which may be regarded as more 'operator-driven', was shown to be satisfactory.

As for the role of modelling or diagramming directly in designing for manual back-up, we are unaware of any studies that have addressed this issue directly. One comment which could be made, however, is that much previous modelling/diagramming work has subscribed to a technical philosophy of functional allocation. That is, it has most often been presumed that manual functions are less effective than automated functions, particularly under conditions of time-stress. This however may not always be the case (Wiener and Curry, 1980). For example, Kurke's (1961) operational sequence diagram illustrated the superiority of a computerized ship-avoidance system. Generally, there is a lack of modelling/diagramming studies that have investigated whether manual back-up is actually possible during conditions of equipment failure, and that have forecast subsequent system performance.

6.7 Effects of Unreliable Data

During discussions of decision-making aids (Section 6.3), hypothesis and option generation (Section 5.4) and decision-making vs. stereotyped behaviour (Section 6.5), a common theme has been that the lack of structure in many real-life tactical situations prevents a straightforward analysis. Similarly, many real-life problem-solving tasks are characterized by the decision-maker being required to act on data of low fidelity i.e., the diagnostic value of the data is small. This latter aspect has the greatest effect upon the hypothesis evaluation stage of decision-making (see Figure 29). In military context, the typical case of

hypothesis evaluation under conditions of unreliable data is that of intelligence analysis.

A number of studies have concentrated upon the hypothesis evaluation stage of tactical military decision-making, as that stage is relatively amenable to a mathematical interpretation. Most studies have utilised some form of prototype, and are reviewed comprehensively in Meister (1976). In this report we shall concentrate on one study alone, namely, that of Howell and Gettys (1968). Their simulation was designed to assess the effects of various factors upon C^2 system performance, with particular emphasis upon the threat evaluation function. The task was sufficiently well-structured to allow a Bayesian interpretation and, therefore, the implementation of an automated decision aid. Both manual and semi-automatic conditions were considered. It was found that the Bayesian aid was of generally high value but was particularly effective when the probabilities of the input data were low. The authors attributed that effect to the inability of operators to conceive of the system in indeterminate terms, i.e., the operators tended to form a hypothesis about the state of the threat and then act as if the probability of that hypothesis was 100%.

Meister (1976) has extended this result by claiming that computer aiding (i.e., Bayesian aiding) is of greatest value "when the data the system must operate on are contaminated, incomplete, nonindependent or otherwise faulty" (p.221). In addition, Vittal et al (1981) have recommended several AI techniques for overcoming the problems caused by misleading information. Those techniques include knowledge representation and inference capabilities, graphical displays and fast-time simulation. Generally, it would appear that the implications of misleading information need to be cross-checked by a number of methods in order that a confident evaluation may be made of the impact on system performance.

6.8 Operator Strategy and Design

Operator strategies may be contrasted with specified operational procedures. One role of the human factors team during later stages of design is to formulate the most efficient operating procedures. Training programs may then be devised and implemented in anticipation of the system becoming operational. A second function of procedural design is that it may compensate for poor engineering design. That is, whilst the system constrains the available procedures somewhat during operation, a choice of procedures may still be possible and attention to these may improve system performance, e.g., Blum, Callahan, Cherry, Kleist, Touma and Witus (1980).

In many instances operator strategies result in informal modifications to the specified procedures. From a human factors engineering perspective, possibly the most important strategies are those that are employed to cope with excess workload. For example, operators may perform a number of tasks in parallel rather than sequentially if the opportunity arises. Such strategies may have the desirable effect of increasing system performance above that which was predicted but, in other circumstances, operator strategies may be inefficient or may conflict with the performance of other tasks.

Obviously, it would be useful to be able to predict such strategies through some form of system model. However, it is doubtful whether the state-of-the-art is sufficient to permit such forecasts. Most

network models, in particular, presume a fixed task sequence (if only in a stochastic sense) and cannot predict tasks or task combinations that were not revealed in the original systems analysis (for further discussion of this issue, see Section 4.6.4). That is, models tend to neglect the emergent properties of systems. Baron et al (1980) have speculated that the PROCRU model, which represents a synthesis of the top-down and bottom-up approaches, may overcome the latter deficiency, although the evidence is yet to be seen.

The preceding discussion of operator strategy has emphasized procedures and task combinations. However, there is a second sense in which the expression 'operator strategy' is used, and that refers to the heuristics employed by the operators during problem-solving tasks, such as tactical C^2. Decision-making may be a peculiarly individualistic affair, and there is evidence that a large variety of styles exist (Meister, 1976).

Most applied studies, e.g., Gagliardi et al (1965), have concentrated on the identification of inefficient heuristics in order to make recommendations for automation, i.e., for decision-making aids (Tversky and Kahneman (1974) have analysed many such heuristics). This provides a basis for a design perspective relating to automation. Additionally designers should ensure that efficient heuristics should possibly be supported by the design of the system. For example, Nickerson et al (1981) believe that many (non-specialist) computer operators and process controllers develop 'mental models' of the functioning of the system in order to compensate for the lack of 'comprehensible physical reality' of computerized systems. The system design (including the software component especially) may either enhance or conflict with that mental model. It has therefore been argued that the manner of information presentation at least should be congruent with the operator's conceptual model (Hollnagel and Woods, 1983). This recommendation provides a variation on the familiar theme that the requirements and limitations of users should be discovered before system design commences.

While the existence of different operator strategies makes the prediction of system performance difficult, complications may also arise when the operational system is being evaluated. Strategies tend to be developed with increasing experience of a system, so it is possible that the characteristics of human performance using a new system may change progressively in the early stages of operation. An initial evaluation of the operational performance strategies may thus differ significantly from that observed after experience has been obtained. (One would expect that operator strategies develop so as to improve system performance with time). Ideally, therefore, personnel should be given the opportunity to develop various skills and strategies before a final systems evaluation is made.

6.9 Individual Differences and Systems Design

It is an axiom of human factors engineering that a good system design should cater for the majority of the user population. This is particularly recognised in the area of workspace layouts, for example. Anthropometric data regarding the dimensions of various user groups exist for the Australian military population and allow specifications to be made about the ease of reach of controls, table heights, etc.

While variations in physique are an important design consideration, individual differences in skill have more relevance to the functional aspects of C^2 systems. Systems should not only accommodate users physically, but a high proportion of potential users need to be able to operate the system effectively at the cognitive level. One role of human factors during system design is therefore to access the skill level of the proposed user population and to translate this information into design constraints. It should be noted that these recommendations do not necessarily exclude the possibility of training or specialist selection as a means of ensuring system effectiveness. However, personnel costs are of increasing concern to the military and there has arisen a corresponding concern that operators should be able to transfer from some current system to a proposed system with the minimum of re-adjustment. That is, the inter-operability of systems has become a human factors design consideration.

System inter-operability is of special relevance to the military because of some personnel practices which emphasize job-rotation across different systems (Smith, 1980; Parrish et al, 1981a). Further, the advent of interactive computer systems has heightened awareness of the issue, because these systems are commonly multi-purpose and must cater for a wide variety of user groups. In particular, there tends to be large variability in the operating skills of military users (Ramsey and Atwood, 1979). Three generic user groups are commonly distinguished, namely, naive users, managers/commanders and technical specialists. Naive users and managers/commanders have only a relatively under-developed operating skill; on the other hand, technicians/specialists are just the opposite and may, in fact, design or modify their own systems. Those system models that take account of individual differences in speed and/or accuracy of task performance may be used to assist forecasts of training requirements. The Siegel and Wolf model, SAINT, and HOS all contain parameters that represent individual differences, although the Siegel and Wolf model is the only one which has been applied to an analysis of training requirements. Siegel et al (1978) have demonstrated how their model may be used to trade-off the improvement that may be expected in system performance against greater operator ability. Siegel et al (1976a) did, in fact, recommend that a training programme might alleviate some of the problems identified by their model in the AN/SQS-26 and AN/SQR-10 system. On the other hand, Siegel et al (1976b) suggested that training would be an inefficient means of improving human performance in the AN/SQS-26, LAMPS and AN/SQR-19 system, and that fundamental system re-design was necessary.

A special category of skilled performance that has received attention by human factors engineers recently is decision-making. Accordingly, the topic of individual differences in decision-making style has emerged as a design issue. The major focus of this work has been on the design of _adaptive_ decision aids for C^2, ie, on the design of aids which may complement the styles of various individuals. The Perceptronics organisation in the U.S. has been a significant contributor to this field, represented by Steeb, Artof, Crooks & Weltman (1975); Samet, & Davis (1977); Steeb et al (1977); Leal et al (1978) and Chu et al (1982). _Although these studies have investigated diverse systems through_ prototypes of remotely piloted vehicle control, anit-submarine warfare and advanced aircraft, a conceptual link has been the philosophy that operators should be provided with information upon which they place relatively great personal weight. The decision-aids thus automate the

information selection function to some extent by 'capturing' individual operator strategies. The details of that process are beyond the scope of this report, but the algorithm has most frequently been based on multi-attribute utility theory or pattern recognition techniques.

6.10 Systems Flexibility

An appeal for flexibility of system design is frequently made with human factors considerations in mind. For example, as discussed in Sections 6.2 and 6.6, the possibility of a dynamic or flexible allocation of function may have the desirable effects of allowing operators to regulate their own work load and of increasing system survivability through manual backup.

Flexibility tends to receive a reasonable degree of attention during the design of interactive computer systems. As discussed in Section 6.9, a feature of these systems is that they often must cater for a variety of user groups possessing of different skills. Consequently, some are more adept that others at using the system to assist them with task performance, eg, through retrieval of information from a data-base.

Managers/commanders also frequently organize to have a specialist available in case they encounter difficulties on the system (Carley, 1967). This may have the effect of causing a shift in the power structure of the command. A frequently suggested solution, there fore, is that the mode of operation of C^2 systems should be flexible in order to accommodate various abilities. This can obviate the need for a "standby" specialist.

In particular, the type of interactive dialogue has received significant attention. Naive operators generally require computer-initiated dialogue, ie, one in which the computer generates queries to which the user must reply (Ramsey and Atwood, 1979). The form of this response is also often structured, eg, through selection from a menu. However, experienced users tire quickly of such systems and prefer user-initiated dialogue. Mixed-initiative or variable-initiative dialogues have therefore been advocated as a means of catering for the majority of users.

A more profound aspect of flexibility than the accommodation of individual differences is the ability of systems to handle changes in their goals or purpose over time. In other words, the growth potential of systems may be important (Carley, 1967). In the field of software design, software reconfiguration, or maintenance (Smith, 1980) is a major concern. The design of this type of flexibility is by no means a straight-forward affair because, for example, it is possible that highly flexible systems may lose power, due to their wide domain of application. Alternatively, it may be possible to build systems that are both flexible and powerful, but at the cost of making them too complex for non-specialist users. A discussion of the precise relationship between software flexibility, complexity and power is beyond the terms of this report but good reviews may be found in Ramsey and Atwood (1979) and Nickerson et al (1981).

6.11 Voice vs. Non-Voice Communication

Within the topic of vocal vs. non-vocal modes of communication, it is legitimate to refer both to person-person communication and person-machine (ie, computer) communication. As regards the latter, humans conventionally interact with computer systems in a visual mode, ie, through a keyboard and CRT. However the use of automated speech recognition is a possibility. Speech generation by the computer to the human shows significant promise. Whilst much research is currently underway in this field, few design guidelines yet exist (Smith, 1980).

Person-person communication, on the other hand, appears to be one of the human factors issues during design that has been considered in detail experimentally. A number of studies have investigated the merits of various communication modes during simulated tactical tasks, ie, through the use of prototypes. The major focus has been on comparisons of visual and verbal modes of communication. Two studies were especially illustrative:

(a) Howell and Gettys (1968) were largely concerned with the applicability of a Bayesian decision-aid during a simulated threat evaluation tasks. The task was a group effort and involved communication between those who were responsible for data-relay and those who performed the actual decisions. The general conclusion was that a vocal mode of communication was not superior and, in fact, showed some tendency to congest the communications channel. An additional feature of the study was that operators had to deal with both probabilistic and all-or-none intelligence data. The latter condition degraded performance, and was not alleviated by vocal communication.

(b) Chapanis et al (1972) were concerned with studying communication modes that can be used in generalized problem-solving task. The vocal link was shown to be superior to an operator-written or typed link with regard to problem solution time. An analysis of operator performance showed that both receiving and transmitting times of messages were shorter when vocal communication was allowed. Such a result might have been expected. However, it would appear that the task conditions were such that the quality or precision of information transmission was not a highly significant factor in team performance. Otherwise, the value of printed communication may have increased.

From the preceding studies, it may be appreciated that any discussion of the relative merits of vocal and non-vocal communication should take into account the particular task properties involved. Additionally, if an auditory mode of communication is to be used, it is reasonable to ask whether speech or some coded form of communication is preferable in certain situations. Many of the relevant principles may be found in Woodson (1981). Data on more molecular design problems which are related to the above issues (such as tolerable signal: noise ratios) may also be found in Woodson (1981).

6.12 Graphical v. Textual Displays

The merits of pictorial and textual modes of information presentation constitute another issue that is relatively concrete. In fact, this issue partially qualifies as a 'human-machine interface' design problem, because it impinges upon the basic perceptual capabilities of

operators. For example, ease of recognition of various displays is a typical interface problem which would arise relatively late in the development cycle, and for which some recommendations exist. On the other hand, the issue is not quite so straight forward because it is the information transmission properties of graphical and textual displays that are primarily being compared, thus requiring some consideration of the human's cognitive abilities.

Textual displays have undoubtedly been favoured by both researchers and designers in the past. Within that topic, a number of specifications exist regarding line spacing, colour coding, etc. By contrast, design guidelines for graphical displays are deficient (Smith 1980). Similarly, the research concerning the relative merits of graphical and textual displays is fragmented and difficult to integrate into general principles (Ramsey and Atwood, 1979). The issue has been further confused by the emergence of interactive graphics, in which the operator may not only call up a particular display page, but may modify the format of that page as well. Well-designed, interactive graphics may be very usable (Bennett, 1978). This topic tends to relate directly to the field of artificial intelligence (Rebane, Walsh and Moses, 1979; Bechtel, 1981).

There are at least three factors which may govern the choice of textual or graphical displays (Ramsey & Atwood, 1979). The first principle is almost tautological, namely, the type of information which one wishes to transmit is an important consideration. Graphical displays have been shown to be superior in tasks that involve the processing of geographical or spatial information. Bechtel (1981) has emphasised the importance of these tasks within C^2.

Another principle, although not uniformly supported by research, is that graphical displays tend to be interpreted faster than textual displays (Tullis, 1981) but are inferior if detailed information processing is required. Correspondingly, if the operator is required to commit much information to memory, textual displays may be preferable. These principles are by no means immutable, because there is the problem that many of the conclusions which have been drawn from past research are extremely dependent on the tasks used. Ramsey and Atwood (1979) and Tullis (1981) give further details of the appropriate studies.

Various methods exist for evaluating the human engineering aspects of visual display, although none specifically address the graphical/textual issue. At preliminary stages of design, these methods rely heavily on the use of expert judgement. Both the Analytic Profile System (APS) of Siegel et al (1975) and the Decision Quality Metric (DQM) of Landis et al (1967) utilise expert ratings in order to assess the informational properties of visual displays (for further details see Section 5.2.2). Both methods have presumably been developed for textual displays, but in principle could be modified for graphics.

Tainsh's (1983) job process charts were used for probably the most specific application related to this issue to date. These diagrams were developed in order to analyse the tasks of British Naval tacticians, particularly those concerned with scenario generation. That study demonstrated how graphical and dialogue-based tasks might be compared, although no firm conclusions were drawn.

6.13 Team Structures

A reasonable amount of applied research exists regarding the effects of team structure upon system performance, although there are relatively few studies that have a direct military application. Team structure has conventionally been regarded as an organisational development problem. In principle this issue could be addressed during design, particularly during the later phases.

The concept of vertical v. horizontal structures is fundamental to team behaviour (Hallam and Stammers, 1979). Vertical structures are those in which individuals are assigned particular functions and must relay the results of their work to others. In contrast, horizontal structures are characterized by the fact that individuals share the total task. The latter organisation corresponds to one in which individual autonomy is relatively high, because performance is not so much driven by the demands of others. Hallam and Stammers (1979) have found (using a C^2 prototype) that both types of structure have their merits, depending upon such variables as task complexity, information processing demands, and response requirements. However, it was also concluded that the C^2 function would frequently benefit from a horizontal structure, in contrast to what is traditional military practice. Wesson, Hayes-Roth, Burge, Stasz and Sunshine (1981) made a similar finding using a simulated intelligence evaluation environment, i.e. a hierarchical committee was considered to be inferior to a uni-level organisation. Many other studies of the effects of team structure upon system performance may be found in the reference lists of these reports.

Team structure also tends to impinge upon some social issues. In the design of human-machine systems, the interaction between humans may be a significant variable, that, if neglected, may have negative results (Cohen and Turney, 1972). For example, it may be that individualized VDU's are effective in a performance sense, but are rejected by the operators because of the loss of social contact. Further details of social factors within computerized systems may be found in Bjorn-Anderson (1978).

Most studies of team structure, or of the allocation of tasks between operators, have used prototypes, e.g. Hallam and Stammers (1979), Jorgensen and Strub (1979), and Wesson et al (1981). Studies by Lindquist et al (1971a) and Siegel et al (1976b) are distinctive in that the optimal workload for individual crew members was determined through the use of diagramming and modelling, respectively.

6.14 Training

Training has not conventionally been regarded as a human factors issue because, as discussed in Section 1.1, training of personnel to maintain system effectiveness represents a somewhat antagonistic philosophy to engineering the system in order to achieve the same goal. However, there is increasing concern within the military that training matters should be revised along with design procedures (Thorndyke and Weiner, 1980; Baum, Modrick and Hollingsworth, 1982; Gardner, 1982).

The introduction of computerized systems has been a significant stimulus to increased concern for training. On-line tutoring, or embedded training, is now a possibility, yet few priniciples exist (Nickerson et

al, 1981). Simulators such as skills trainers (Maltback, 1980) are proving useful as research tools in this area beyond their specific training capability.

It should be recognized that the issues of communication (Section 6.11), team structure (Section 6.13) and training are all intimately linked, which makes discussion of any issue in isolation difficult. For example, Baum et al (1982) claim that, while the training of individuals is relatively well-developed, principles for team training are deficient and require research. Siegel and Federman (1973), in attempting to train helicopter ASW teams, focussed on communication performance. O'Reilly and Roberts (1977), and Hallam and Stammers (1979) have also emphasized that the effectiveness of certain team structures may be mediated by intra-team communication. Meister (1976) provides a good summary of such issues.

From a design or procurement viewpoint, the training issue becomes one of forecasting the necessary skill-level of the potential operators. Ideally, the skill required should not be greater than that which is available within the current personnel resource (Smith, 1980). However, if a discrepancy exists, then a training programme and/or specialist selection procedures may be suggested (with another alternative being system re-design).

As discussed in Sections 2.11 and 3.7, most modelling/diagramming techniques have failed to address cognitive behaviour to a sufficient extent. The possibility of forecasting training requirements at the cognitive level for many tasks within computer systems is therefore limited.

Aside from the function of forecasting the skill level required to operate a system, modelling/diagramming techniques have other uses that have an impact on training issues. These techniques permit a rather specialized means of task analysis (or, more precisely, permit a specialized representation of the data obtained from task analysis). Such analyses are a prerequisite for the institution of a training schedule (Silvern, 1970; RAN School of Training Technology, 1978).

6.15 Personnel Estimates

As with the issues of training (Section 6.14) and system inter-operability (Section 6.9), personnel estimates are of constant importance during system design. For economic reasons, systems cannot rely on excessive numbers of personnel for operation, and must be designed within certain constraints. As discussed in Section 1.2.2, one likely human factors issue is the need to forecast personnel requirements from a conceptual design. This information may and should provide a check on the development contract. Further, automation is not necessarily the best means of reducing system cost, as maintenance factors tend to increase and system flexibility is often reduced (see Section 6.2.2).

As for methods of personnel forecasting, the models reviewed in this report have only indirect use. The main reason is that the majority accommodate single operator performance alone. As discussed in Section 3.2 it is theoretically possible to predict the performance of groups by extrapolation from performance of individuals, but many difficulties arise. Generally, team interactions are a confounding factor. As discussed by Baum et al (1982), our theoretical knowledge of

group processes within military systems is far from adequate, which suggests that the construction of suitable models requires further research.

As noted in Section 3.1, we have also not considered many models which are solely concerned with the allocation of a given personnel resource to a system. For reasons of convenience, we have focussed on models which evaluate the human factors engineering adequacy of a system by forecasting operability. One exception, however, is the group model of Siegel and Wolf which has been developed for the U.S. Navy (see Section 4.2.2). The study of Siegel et al (1978) illustrates how different manpower policies may be traded off against expected system performance.

Expert judgement has also figured as a means of making personnel forecasts, in contrast to formal modelling. A brief description of the former approach may be found in Section 5.2.4.

7. RECOMMENDATIONS

7.1 Introduction

The major aim of this report has been to survey various methods that are useful for applying human factors principles and analyses during system design or procurement, with particular reference to RAN combat data systems. Accordingly, a literature search(that principally considered design studies performed under contract to the U.S. military) was undertaken with this goal in mind. As a result, the general approach of this report has been descriptive, in the sense that the report reviews what methods human factors workers have used in the past, without attempting the development of new methods. One function of this report, therefore, is to delineate the state-of-the-art in applied human factors methods. As discussed in the introduction (Section 1.3), the methods to be reviewed were restricted for reasons of brevity and salience primarily to those which are applicable at relatively early stages of design.

In the course of reviewing these methods (especially human performance diagraming and modelling techniques), it was possible to make a comparative evaluation. This evaluation was based on empirical data when such were available, and on logical analysis where appropriate. The survey and analysis showed that no one method may be regarded as generally superior, but that different methods have different purposes and characteristics. The evaluation of those methods may be found in the summaries of the diagramming and modelling chapters (Sections 2.11 and 4.7, respectively).

The purpose of this chapter is to augment the descriptive aspects of this report with some prescriptive recommendations. Certain general principles have emerged from the wide range of literature that has been surveyed. It is our intention to communicate those principles for the benefit of system designers and procurers alike.

The recommendations basically fall into two categories, what have been termed 'research' and 'managerial'. The research recommendations have resulted from our analysis and perception of certain deficiencies within the literature, and constitute a program of investigation. The selection may be biassed somewhat because of our inability to obtain some information that is not present in the open literature, but the consensus of specialists also suggests that many areas lack adequate research. The managerial category, alternatively, contains recommendations that are believed to constitute good human factors practice during design. Many of the recommendations that have been selected for emphasis represent contemporary thinking and are not discussed in traditional human factrors engineering textbooks.

From a procurement viewpoint, therefore, these latter recommendations may be interpreted as forming a basis for new contractual guidelines.

7.2 Managerial Recommendations

7.2.1 Introduction

It is considered that, despite the existing gaps in human factors expertise which have been described in the preceding section, improved

management practices during design will significantly enhance the quality of the human factors input. That is, the most immediate initiatives that are required are managerial rather than technological (see Goodbody and Monteleon, 1976).

The recommendations in the report will be written from a procurement viewpoint, i.e. they provide contractual guidelines. An attempt has not been made to write a complete contractual specification for a system development project. Rather, recommendations have been emphasized that are not considered by traditional human factors engineering textbooks. In particular, it should be noted that design practices at early stages of development are critical to project success, and the recommendations accordingly are mainly associated with that phase. The order in which the recommendations are presented should not be taken to be a listing according to importance. However, some of these are more general than others and thus are relevant to a wider range of situations.

In some instances, little attempt has been made to provide detailed support for these recommendations. The support may be found in specific sections of this report, and those sections have been indicated. However, many of the general recommendations are an outgrowth of the report as a whole and are discussed in some detail here.

7.2.2 Requirements definition

It is a characteristic of the systems approach to design that the requirements or the goals of the system should initially be specified in order that the means of achieving those goals may be designed. Within computer systems at least, it is widely recognised that the potential users of the system should assist in the formulation of those goals (Ramsey and Atwood, 1979; Smith, 1980). Documentation should be required to ensure that the potential users of a system have been consulted about their requirements and attributes before system design continues. As noted by Garrison (1980), inadequate documentation at the planning stage of design is a frequent cause of management information system failure.

The precise method of consulting users may be variable according to the circumstances involved. For example, it may be sufficient to consult a representative sample of users if many perform the same tasks. Similarly, it may not be necessary to consult users exhaustively if experience with a previous system provides guidelines about their requirements. Some of the users' attributes and requirements can also be discovered by consulting human factors engineering textbooks (e.g. McCormick and Sanders, 1981) and handbooks (e.g. Woodson, 1981).

At more detailed stages of design, user consultation may also necessitate that the tasks of the potential users analysed. This analysis should ensure that the capabilities of the proposed system users have not been exceeded. User involvement should preferably be as direct as possible because, if the 'users' engaged in system design are actually experts, they will be able to interact with the system whether or not it is designed optimally (Nickerson & Pew, 1977). Further discussion of user involvement may be found in Appendix A.

Recommendation 1: **User requirements should be formally investigated at the planning stage of system design and a document should summarize the requirement.**

7.2.3 Use of scenarios

A major problem with many system development projects is that the criteria of human performance have not been specified in detail. That is, while the overall goals of the system may be specified in the development contract, the level of human performance which is required to meet these goals is often uncertain. It is then difficult to decide what constitutes an acceptable or unacceptable level of human performance because the relationship with systems effectiveness is unclear.

It can be argued that the use of scenarios at the planning stages of design can alleviate this problem. The scenario begins as a description of a typical mission which will be undertaken by the system. The description then becomes more detailed as quantitative values are inserted, e.g. travelling distances, mission time, etc. After the functions of human and machine have been differentiated (see Sections 1.2.2 and 6.2), it should become possible to ascertain the constraints within which human performance will occur. The time-limits for various functions and tasks should at least be identified; and the information required for that performance should be described.

Naturally, the scenario does not have to include every aspect of the system mission. It should include those human functions which are critical to mission success. An estimate of the effects of human failure within crucial tasks upon system performance should be attempted.

Recommendation 2: **Criteria of human performance need to be specified at planning stages of design. The contractor should respond to the global system requirements specified in the development contract with a scenario that delineates the criteria which human performance must meet in order to maintain system effectiveness.**

7.2.4 Form of evaluation

Generally, the onus should be with the contractor to demonstrate that the system is well-engineered from a human factors perspective. It is not sufficient for the contractor to state that various military specifications have been adhered to, as these are frequently too general to be very useful. Rather, the contractor should provide a formal evaluation of the system at intervals during system development in order to demonstrate its human factors engineering adequacy.

Evaluation should preferably involve a comparison between projected human performance and the criteria of performance that have been derived from the scenario. A weaker, but still desirable form of evaluation is one in which human performance within alternative system designs is compared and is shown to be superior under one design.

The results of the evaluation should be documented in a form that can be examined in detail. Thus, if multiple contractors offer competing systems, the documentation should facilitate comparison (see Section 4.7.2). Alternatively, if various system concepts have been entertained

by the one contractor, the criteria for selecting one should be apparent. If a system concept has been rejected on the grounds of excess costs, any trade-offs with human performance should be described and preferably quantified. In effect, all evaluative work should be documented as if the contractor were teaching a third party how to do it (Smith, 1980).

> **Recommendation 3: The contractor should formalize all methods of human factors evaluation and then document the results in a clear fashion..**

7.2.5 Use of expert opinion

Following on from Recommendation 3, if expert opinion is used as a means of evaluating or verifying a design, that evaluation should be open to scrutiny. This suggests once again that the method of evaluation should be documented precisely, along with the criteria upon which a particular design decision has been made.

Generally, expert opinion should be structured as much as possible. Binary decisions of whether a system, or feature of a system, is acceptable or not are of little value. The dimensions of the system upon which the judgement rests should be defined. That is, while many evaluative techniques are founded upon the use of expert judgement, some are more effective and valid than others. The efficiency of human factors checklists can be questioned and should be avoided, or at least used only as a screening device. For a fuller discussion of this issue, see Section 5.1 and 5.2.

> **Recommendation 4: Expert opinion as a means of systems evaluation should be structured and well-documented.**

7.2.6 Use of diagrams and models

A central theme of this report has been that human performance diagrams and models permit early human factors input to the system design process. While diagrams are commonly associated with a graphical description of performance and models are associated with digital behavioural simulation, both techniques may be classed as 'models' in a broad sense because they provide an abstract representation of system functioning. By this definition, therefore, scenarios also qualify as system models. All these models may be used for purely descriptive purposes or, alternatively, may be used for performance forecasting if quantitative data are inserted.

During scenario development, particular attention should be reserved for those human functions which are subject to significant time-contraints. Such contraints lead to stress, and increase the probability of human error. A good 'rule-of-thumb' is that of Jones & Wingert (1969), which states that the time required to perform a function should be no more than two-thirds of the time available. If not, system re-design is recommended.

The use of the time-line scenario for the purpose just described constitutes a crude form of modelling. Performance time for functions may be estimated in a global fashion, or may be obtained by summating performance times for individual tasks which comprise that function.

Similarly, mission duration estimates may be obtained by summating function times (presuming these occur serially). Despite the methodological difficulties with such simple analytic models (see Section 3.2.1), their application is advocated as a means of performing systems evaluation at an early design phase, and particularly as a means of checking the human-machine allocation (see Section 1.2.2 and 6.2).

At this stage of design, some attempt should also be made to analyze the information flow within the system. The information which is necessary for the performance of each human function should be tabulated (for example, in an aircraft approach-to-landing, one at least requires information about speed, altitude, obstacles, weather, etc.). The use of operational sequence diagrams is supported (see Section 2.6) for depicting this analysis: As noted by Parks and Springer (1976), these diagrams formalize what is often implicit in the scenario and functional analysis, and provide a good overview of the procedures within the system. The diagrams also provide a gross check on the information flow within the system, i.e. if a discrepancy exists between the information which is required and that which is available, system re-design is suggested.

Modelling is also appropriate (and has probably been used extensively) at detailed stages of design, e.g. in the design of controls and displays. While it is desirable that the contractor should apply such techniques, no recommendations about later phases of development are made in this report. Attention has instead been concentrated on models that may be derived in a relatively immediate fashion from the mission scenario.

System modelling procedures can also be carried out by the procurement team. The procurer should have access to the simple analytic models and operational sequence diagrams of the developer during the initial phases of design; or alternatively, these models could be derived in collaboration. Possession of these models could allow the procurer to take a more active role in the design project than otherwise. The models facilitate the ability of the procurer to evaluate the human engineering adequacy of the conceptual design and increase the procurer's ability to make design inputs, particularly as regards the human-machine functional allocation (see Sections 1.2.2 and 6.2). Other, less prescriptive modelling recommendations may be found in Section 3.7.

> **Recommendation 5:** Simple time-based analytic models and operational sequence diagrams should be derived from the mission scenario. The time constraints of functions and information requirements should at least be analyzed and documented as a formal contract report.

7.2.7 Document stores

In our discussion of models (Recommendation 5), it was proposed that contractors should grant the system procurer access to models and drawings of the conceptual system in order that the procurer could take a more active part in design. That philosophy may be extended by recommending that, to the extent possible, the procurer should actually have a document store of related systems that are in operation. The documents would include scenarios, drawings, models, etc, which represent the functioning of such systems. Naturally, these documents associated with the conceptual system could be modified after the system became operational if their projections were shown to be inaccurate.

The advantages of such a document store would be two-fold, although both advantages are related. First, if system re-design was contemplated, much time spent on forecasting performance could be saved because basic (valid) models of the original system would already be in existence. That is, design redundancy would be avoided. Secondly, the document store would ensure that designers and procurers alike learnt from previous systems. This is currently an uncertain process (U.S. General Accounting Office, 1981).

Functional flow diagrams are particuarly amenable to storage and hence future reference. Depending on the system involved, many functions remain the same even after system re-design, because the most common design innovation is re-allocation of function between humans and machines (see Section 1.2.2 and 6.2) rather than a change in the functions themselves. The retention of molar functional flow diagrams (i.e. those which do not differentiate human or machine) would therefore require little successive modification to the document store. For example, Parks & Springer (1976) have emphasised the similarity of functions between aircraft systems at least. Retention of functional flow diagrams is seen as a means of speeding up human factors evaluation at early stages of system re-design.

While the general view that one should learn from previous systems is laudable, there is a danger that re-design may be less innovative due to inappropriate reliance on past experience. Past experience should be useful insofar as it provides a conceptual framework of the system functioning. However, it is less desirable for previous technical design solutions to continue to exert an influence. Once again, this suggests that the retention of abstract models of the system (such as molar functional flow diagrams) should be given priority.

> Recommendation 6:   A document store which includes abstract models of the functioning of all systems in operation would facilitate system re-design.

7.2.8 Computer-aided design

Computer-aided design is becoming increasingly prominent, and deserves to be mentioned in this report. The actual use of the computer per se is not the most salient aspect, rather, the essence of the technique is that design must necessarily be performed in a systematic and top-down manner. That is, a systems approach to design is required. Two interesting techniques from a human factors viewpoint have been identified: CAFES and SADT.

The Computer Aided Function Allocation and Evaluation System (CAFES) (Parks & Springer, 1976) has been developed by Boeing for the U.S. Naval Air Development Center. The technique consists of the successive application of a number of models to evaluate the system. Many of these models have already been discussed individually, including functional flow diagrams, time-lines, operational sequence diagrams and HOS. The value of CAFES it that is integrates all these techniques and orders the method of evaluation into a logical hierarchy.

In many ways, the technique provides a model of how hardware development should be carried out. (The technique may be less relevant to

computer system development, such as for C^2). Starting with requirements definition, CAFES partially automates the evaluation of functions, functional allocation, interface design and workspace layout according to certain criteria. It facilitates forecasts of manpower and training requirements. Like all models, the technique requires comprehensive input data.

The Structured Analysis and Design Technique (SADT) (Ross & Schoman, 1977), alternatively, has its main application within software development. Commencing with requirements definition, the technique partially automates the se.ection of system functions which will achieve those system goals. In principle, the technique may be applied at later stages of design and is not restricted to software development.

A significant value of SADT is that it provides a graphical means of decomposing system functions. Davis (1982) has illustrated how this approach may be used to design human-machine interfaces, in conjunction with the SAINT simulation language (see Sections 2.7 and 3.2.5).

> **Recommendation 7: Computer-aided design techniques are a means of ensuring a systems approach to design. They should be applied where appropriate.**

7.2.9 Use of prototypes and mock-ups

While prototypes and mock-ups have the advantage of allowing a concrete evaluation of the system, verification of the system through these techniques alone is undesirable. Prototypes may only be built when the system is at a reasonably fixed stage of development, thus any re-design at that stage will necessarily be inconvenient and expensive. The danger is, in fact, that prototypes will merely be used to justify a previously made design decision. The use of this technique alone increases the tendency for 'in-house' designs, that often figure in system failures (Garrison, 1980). (A more comprehensive discussion of the merits of modelling and prototype testing may be found in Section 3.1.1.)

Having commenced with a number of negative comments, some solutions may be suggested. The use of reconfigurable prototypes is advocated to ensure that alternative designs are considered. Further, a neutral third party could administer the prototype tests to minimize bias. Topmiller (1981) has made a strong case that a hybrid modelling/prototype evaluative technique may be useful, in order to compensate for those situations in which it is not feasible to model the behaviour of the human. Berson and Crooks (1976) have provided guidelines for the use of prototype tests which could ensure the effectiveness of this technique.

The building of mock-ups and prototypes often necessitates a reasonably large committment of resources. Accordingly, prototypes should possibly be designed to answer more design questions than has conventionally been done. The demonstration of manual back-up possibilities for crucial functions can be regarded as particularly relevant. That is, the ability of the operator to use an automated system in a degraded (or manual) mode should be investigated (see Section 6.6). In addition, the optimum team structure of a system is an issue which could routinely be investigated through a prototype. Finally, while prototypes may be used to evaluate design configurations, the possibility

of testing alternative system procedures within any one design should not
be neglected (see Section 1.2.3).

> **Recommendation 8: Systems should not be evaluated through the
> use of prototypes and mock-ups alone. However, these techniques
> do have the advantage of permitting detailed evaluation, and
> should be utilized in a comprehensive manner.**

7.2.10 Personnel resources

The evaluation of human engineering adequacy has been emphasised
in these managerial recommendations. However, that is not the only issue
which concerns the procurer because it is the cost/effectiveness of the
system in a global sense which is the major consideration. One large cost
of the system, apart from the hardware, is that of the personnel who will
be required to operate it. For this reason, other human factors questions
regarding the numbers and training of the potential system personnel need
to be investigated.

Generally, the contractual specification should describe the
limits of the personnel resource which is available for the operation of
any proposed system. It then becomes the obligation of the contractor to
demonstrate that the system may be operated effectively by that number of
personnel, possessing various levels of skill, etc. In particular, if
automation of part of a system is contemplated, an ideal goal is that the
new system should not require greater skill for operation than the
previous manual system (Smith, 1980). In practice, it is frequently
difficult for the contractor to forecast skill requirements, so a
compromise may be that systems should be shown to be operable by personnel
of the lowest possible training level for a given level of effectiveness.

Both Sawyer et al (1981) and U.S. General Accounting Office
(1981) have proposed that the contractor should be required to document
personnel costs stringently. Some of the latter recommendations are
paraphrased here:

(i)   reduction of manpower and increase in productivity should be
       shown to be constant design goals;

(ii)   tradeoffs between numbers and skill levels of personnel should be
       identified;

(iii) as regards skill, a distinction should be made between on-the-job
       training requirements and a priori qualifications;

(iv)   the availability of personnel for a new system should be
       considered;

(v)   previous staff shortages (through high turnover rates) should be
       identified and resolved within the new system;

(vi)   training lead times should be considered;

(vii) training manuals should be shown to be adequate;

(viii) specific training organisations and programs should be
       identified, and their availability noted;

(ix) specialist skill requirements should be highlighted;

(x) peace-time v. war-time requirements should be distinguished.

Recommendation 9: The characteristics of the available personnel resource, namely numbers and training, should be a contractual specification that bounds the design. The contractor should provide documentation to demonstrate that these limits will not be exceeded.

7.2.11 Subsystem integration

It is a feature of the Australian procurement cycle that various sub-systems are often purchased from overseas contractors, and are then combined to form the total system. Such a method may be described as a 'building block' approach to design (Clapp and Hazle, 1978), and may have a number of beneficial aspects. In particular, the approach avoids the situation whereby a contractor is inundated with specifications and must then produce a total system which conforms to those specifications by whatever means are available. The approach allows incremental experience to be gained with a system before a new sub-system is added, and is congruent with the evolutionary approach to design.

On the negative side, the building block approach to design conflicts somewhat with the systems approach, because design may not necessarily be deriven by a total systems concept. Consequently, difficulties may be experienced when attempting to integrate a new sub-system with the old, due to the fact that the systems may have been designed according to different criteria. Further, there is the possibility that, while the various systems may function adequately together in a technical sense, the total operability of the system may be low.

The solution appears to be that, if a building block approach to design is taken, the systems approach should also be retained. New sub-systems should not be tested and evaluated in isolation, rather, the environment in which those sub-systems will have to perform should also be considered. From a human factors viewpoint, the operability of various combined systems should be a major concern.

Recommendation 10: If a building-block approach to design is followed, sub-system integration should be demonstrated.

7.3 Research Recommendations

Generally, the research proposals presented here may be further subdivided into two categories, based upon the type of deficiency that has been observed in the literature. The first category concerns lack of data for various design issues, which has been mentioned previously in Chapter 5. The second type of deficiency observed results from the lack of refinement or availability of certain human factors techniques. The following discussion will treat both categories separately.

For a broader discussion of the problems facing human factors in systems generally and an indication of productive avenues of research, both Topmiller (1981) and Meister (1982b) should be consulted.

7.3.1 Data needs

In the discussion of selected human factors design issues in Chapter 5, it is difficult in many cases to provide firm guidelines because of either a lack of research, or the equivocal nature of the findings. Many issues, particularly those arising within the more recent developments in computer science, have not been sufficiently researched for design guidelines to be formed. Many of these systems (such as artificial intelligence systems) are designed without drawing directly on past experience or a standard body of data. Even within the more conventional areas of system design, a lack of specific design principles and specifications often results in the onus for consideration of human factors issues lying with the contractor, which is to some extent undesirable.

More particularly, within the issues embraced by this report, research needs were identified in the areas of:

(i)   Manual back-up (Section 6.6). Few systems are designed under the expectation that they may still be operated during conditions of computer failure. For example, a prime human factors question would be the ability of a commander to use a decision-aiding system that was performing in a degraded fashion. Research is required to identify the factors which ensure that such systems degrade gracefully from a human performance viewpoint, and to investigate the ability of people to transfer between manual and automated modes of operation. A logical starting point for such research is to ensure that prototype studies include a degraded mode analysis.

(ii) System flexibility (Section 6.10). System flexibility, power and complexity are all important variables which interact. At present, design guidelines can only be generally stated, e.g. "the system should cater for a variety of user groups". Research is needed to improve the specificity of such guidelines.

(iii) Operator strategies (Section 6.8). The effects of operator strategies upon system performance are poorly documented. There is evidence that operators tend to deviate from prescribed routines with experience of a system, but whether this phenomenon is desirable or not is a question for research. Further, there is an increasing recognition that systems should be designed so that they cater for individual strategies, yet the necessary principles are lacking.

(iv) Graphical displays (Section 6.12). There is almost universal agreement that interactive graphics should be beneficial within the C^2 function yet, once again, the design of such systems tends to be left to the discretion of individual contractors. Research is needed to identify the situations in which graphics are preferable to text, and to identify the parameters of such displays that modify performance.

(v)   Team structure (Section 6.13). The optimum allocation of tasks within a team of operators has long been a concern of organisational and training specialists, but there is a growing

awareness that this factor should be considered during design. That is, certain configurations may necessitate a particular form of team structure, that may have unforeseen consequences. In addition, there is some evidence that certain forms of team structure are better suited to certain types of tasks. Research is needed to extend this finding, particularly within a military context, and thus ensure that the team structures implied by a design and required by the task are congruent.

7.3.2 Technique refinement/availability

It is a major tenet of modern human factors policy that effectiveness at preliminary stages of design requires the use of relatively sophisticated analytical techniques, e.g. Meister (1982b). We believe that the most useful techniques are drawings, models and structured expert judgement, and 'have discussed these techniques comprehensively in Sections 2.3, 4 and 5. However, our literature ˉearch has also revealed that many of these techniques are inadequatɛ for answering some important human factors design problems.

The most prominent deficiency of these applied methods is that, generally, they fail to assist performance forecasts of cognitively-based behaviour. Many modelling and diagramming techniques have been developed within the field of industrial engineering and have subsequently been modified for use by human factors engineers. The techniques have correspondingly become more psychological in construct, i.e. they have come to address cognitive behaviour. However, the state-of-the-art in this area falls short of being highly useful to the system designer (Pew et al, 1977, Pew and Baron, 1982).

The immediate consequence of this lack of refinement is that it is especially difficult to forecast system performance when that performance is mediated by the cognitive processes of the human. These processes are most relevant during the design of software generally, and within specialized fields such as decision-aiding systems. Consequently, software guidelines tend to be developed on a trial-and-error basis rather than through analysis. This is in contrast to, for example, guidelines that are available for workspace layouts.

The concern with cognitive behaviour expressed here should not be taken as implying that there is a need for analysis of the cognitive processes of system personnel. The concern is with engineering the informational aspects of system functioning. As the relevant operator behaviour is bounded by the demands of the system, it is reasonable to investigate which type of system design promotes effective information processing and cognitive performance.

Many of the more recent applied methods show promise in this regard. Both job process charts (Tainsh, 1983) and process control diagrams (Drury, 1983) appear to be well-suited to representing the 'internal' strategies of an operator in sufficient detail (see Sections 2.9 and 2.10, respectively). However, these methods have so far been used to capture behaviour retrospectively rather than to predict that behaviour. As emphasised by Nickerson et al (1981), prediction is the essence of design, viz:

"Performance evaluation has always been recognised as an
important component of the system development cycle.
What has been less generally recognised is the importance
of performance prediction. What one would really like to
be able to do is to predict in advance of system
implementation the performance of the equipment, the
user, and the user-machine. Further, one would like to
be able to predict how that performance would depend both
on the characteristics of the system and on the situation
in which it is used. One would especially like to be
able to predict performance in high-demand, stressful,
crisis situations". (p.179-180).

A second deficiency of current models/drawings is that they fail
to address team behaviour. As noted by Nickerson et al (1981), "The state
of model development for large-scale multi-person systems remains crude"
(p.182). Performance of groups is often predicted by extrapolation from
performance of single operators, which is known to be invalid. Some
exceptions do exist, notably some of the models developed by Applied
Psychological Services (Siegel & Wolf, 1981). See Sections 4.2.2 for
further details.

Lack of refinement is probably a major factor that prevents the
more widespread use of human performance models and diagrams. However,
the lack of availability of those techniques may also be an important
consideration. Berson and Crooks (1976) would appear to concur, viz,
"Historically, the discipline of human factors engineering (HFE) has been
handicapped in identifying ... problems because there was not enough
access to drawings and models during the early design stages" (p.5).

This lack of availability was reflected in the literature search,
conducted here in which it was frequently difficult to find contemporary,
detailed examples of the use of models and drawings during design. It was
correspondingly difficult to decide just how useful these techniques have
been and how much human intuition has been involved. Meister (1982c) has
echoed these sentiments: "Useful descriptions of such basic (and
primitive) techniques as time line analysis, operational sequence
diagrams, workload analysis, etc, are almost non-existent. It is even
unclear to what extent these methods are actually used in development and
how" (p.286).

One conculsion that may be drawn from our investigation,
therefore, is that merely making public the use of drawings, models, etc,
may be just as beneficial to the human factors discipline as refinement of
those techniques. Undoubtedly, reasons of both commercial and military
security are powerful motives that prevent such a disclosure, e.g. Zachary
(1980) has referenced five studies concerned with the use of diagramming
during the design of the LAMPS systems: three of the documents are
confidential while one is secret. Tainsh (1983) has also implied that job
process charts may virtually become the blueprint for design within RN
command sub-systems. Apart from illustrating the desirability of human-
centred design, this observation illustrates the potential unavailability
of certain design studies for evaluation.

REFERENCES

AKST, G. Evaluating tactical command and control systems - a three-tiered approach. Alexandria, VA: Center for Naval Analyses, Report No. 358, September, 1982. (AD A122 478).

ALBERTS, D.S. System evaluation and the design and acquisition process. Proceedings of the IEEE International Conference on Cybernetics and Society, Cambridge, MA, October, 1980.

ANDERSON, D.F. The determination of information requirements of a Marine Corps communication manager. Monterey,CA: Naval Postgraduate School, March, 1973. (AD 761 461).

ANDERSON, R.O. A new approach to the specification and evaluation of flying qualities. Wright Patterson Air Force Base, Ohio: Air Force Flight Dynamics Laboratory, Technical Report AFFDL-TR-69-120, 1970.

ASKREN, W.B. Preface. Human Factors, 1975, 17(1), 2-3.

BAINBRIDGE, L., BEISHON, H., HEMMING, J.H. and SPLAINE, M. A study of real-time human decision-making using a pilot simulator. Operational Research Quarterly, 1968, 19, 91-106.

BAKER, J.D. Quantitative modelling of human performance in information systems. Ergonomics, 1970, 13, 645-664.

BARON, S., FEEHRER, C., MURALIDHARAN, R., PEW, R. and HORWITZ, P. An approach to modelling supervisory control of a nuclear power plant. Cambridge, MA: Bolt, Beranek and Newman, Inc. Report No. 481, July, 1982.

BARON, S., KLEINMAN, D.L. and LEVISON, W.H. An optimal control model of human response. Part II: prediction of human performance in a complex task. Automatica, 1970, 6, 371-384.

BARON, S. and LEVISON, W.H. An optimal control methodology for analyzing the effect of display parameters on performance and workload in manual flight control. IEEE Transactions on Systems, Man and Cybernetics, 1975, 5, 446-452.

BARON, S. and LEVISON, W.H. Display analysis with the optimal control model of the human operator. Human Factors, 1977, 19, 437-457.

BARON, S., MURALIDHARAN, R., LANCRAFT, R. and ZACHARIAS, G. PROCRU: a model for analyzing crew procedures in approach to landing. Cambridge, MA: Bolt, Beranek and Newman Inc., Report No. 4373, April, 1980.

BAUM, D.R., MODRICK, J.A. and HOLLINGSWOTH, S.R. Team training for command and control systems: executive summary. Wright Patterson Air Force Base, Ohio: Air Force Human Resources Laboratory Report No. AFHRL-TP-82-11-Vol.5, April, 1982. (AD A113 503).

BECHTEL, R.J. C^2 relevance of selected artificial intelligence systems. San Diego, CA: Naval Ocean Systems Centre, Technical Report TR-706, July, 1981. (AD A105 876).

REFERNCES (CONT.D)

BEISHON, R.J. Problems of task description in process control. In W.T. Singleton, R.S. Easterby and D. Whitfield (Eds.) The human operator in complex systems. London: Taylor and Francis, 1967.

BENNETT, J.L. Incorporating usability into system design: the opportunity for interactive computer graphics. Proceedings of the IEEE International Conference on Cybernetics and Society, Tokyo, November, 1978.

BERSON, B.L. and CROOKS, W.H. Guide for obtaining and analyzing human performance data in a material development project. Woodland Hills, CA: Perceptronics Inc, Report No. PTR-1026-10/76, September, 1976. (AD A-71 196).

BHATTACHARYYA, K.C., PRASAD, U.R. and SARMA, I.G. Cockpit-display design through optimal control theoretic simulation of the pilot-aircraft system. Proceedings of the IEE International Conference on Cybernetics and Society, Washington, DC, October, 1972.

BJORN-ANDERSEN, N. Introduction. In N. Bjorn-Andersen (Ed.) The human side of information processing. Amsterdam: North-Holland, 1978.

BLUM, R.W., CALLAHAN, C.A., CHERRY, W.P., KLEIST, D., TOUMA, G. and WITUS, G. Information management for an automated battlefield command and control system. Ann Arbor, MI: Vector Research, Inc., Report VRI-ARI-3-FR80-1, May, 1980. (AD A109 285).

BONDER, S., CHERRY, W.P. and FARRELL, R.L. A framework for identifying human factors research needs in command and control. Ann Arbor, MI: Vector Research Inc., Report No. VRI-ARI-4-FR80-1, March, 1980. (AD A083-011).

BOWEN, R.J., NICKERSON, R.S., SPOONER, R.L. and TRIGGS, T.J. Army tactical intelligence: an investigation of current methods of production and possible computer-based innovations. Cambridge, MA: Bolt, Beranek and Newman Inc, Report No. 1945, September, 1970.

BRADLEY, J.V. The optimal-pessimal paradox. Human Factors, 1975, 17, 321-327.

BRADLEY, J.V. Paradox lost, paradox regained: Reply from a flagellated straw man. Bulletin of the Psychonomic Society, 1983, 21, 69-72.

BROOKES, C.H.P., GROUSE, P.J., JEFFREY, D.R. and LAWRENCE, M.J. Information systems design. Sydney: Prentice-Hall, 1982.

BURGER, W.J., KNOWLES, W.B. and WULFECK, J.W. Validity of expert judgements of performance time. Human Factors, 1970, 12(5), 503-510.

CARBONELL, J.R., WARD, J.L. and SENDERS, J.W. A queuing model of visual sampling: experimental validation. IEEE Transcations on Man-Machine Systems, 1968, 9(3), 82-87.

CARLEY, J.C. Military command information systems from the user's point of view. In D.E. Walker (Ed.) Information system science and technology. Washington, DC: Thompson Book Co., 1967. (AD 666 336).

CARLSON, D.L. Integration analysis: a proposed integration of test and evaluation techniques for early-on detection of human factors engineering discrepancies. Monterey, CA: Naval Postgraduate School, March, 1983. (AD A127-611).

CHAPANIS, A. Men, machines and models. American Psychologist, 1961, 16, 113-131.

CHAPANIS, A. Research techniques in human engineering. (2nd Ed.) Baltimore: John Hopkins Press, 1962.

CHAPANIS, A. On the allocation of functions between men and machines. Occupational Psychology, 1965, 39, 1-11.

CHAPANIS, A., OCHSMAN, R.B., PARRISH, R.N. and WEEKS, G.D. Studies in interactive communication: 1. The effects of four communication modes on the behaviour of teams during cooperative problem solving. Human Factors, 1972, 14(6), 487-509.

CHRISTENSEN, J.M. Human factors engineering considerations in systems development. In W.T. SINGLETON, J.G. Fox and D. Whitfield (Eds.) Measurement of man at work: an appraisal of physiological and psychological criteria in man-machine systems. London: Taylor and Francis, 1971.

CHU, Y., CHEN, K., CLARK, C. and FREEDY, A. Analysis and modeling of information handling tasks in supervisory control of advanced aircraft. Woodland Hills, CA: Perceptronics Inc, Technical Report PFTR-82-5, May, 1982. (AD A119 067).

CHU, Y. and ROUSE, W.B. Adaptive allocation of decision-making responsibility between human and computer in multitask situations. IEEE Transactions on Systems, Man & Cybernetics, 1979, 9, 769-778.

CLAPP, J.A. and HAZLE, M. Building blocks for C^2 systems. Hanscom Air Force Base, MA; Electronic Systems Division, Report No. ESD-TR-77-360, March, 1978. (AD A052 568).

COHEN, S.L. and TURNEY, J.R. Significance of man-man interactions in the design of automated military systems. Proceedings of the Annual Convention of the American Psychological Association, 1972, 7, 623-624.

CONNELLY, E.M. Research on manned system design using operator measures and criteria (OMAC) data. Vienna, VA: Omenemii Inc, Report No. OTR-62-77-2, July, 1977. (AD A046 464).

COOPER, G.E. and HARPER, R.P. The use of pilot rating in the evaluation of aircraft handling qualities. National Aeronautics and Space Administration, Ames Research Center, CA: NASA TN D-5153, April, 1969.

CORNELL, C.E. Minimising human errors. Space/Aeronautics, 1968, 49, 72-81.

CROLOTTE, A. and SALEH, J. Selecting a decision aid. Proceedings of the IEEE International Conference on Cybernetics & Society, Georgia, October, 1980.

REFERNCES (CONT.D)

CURRY, R.E., KLEINMAN, D.L. and HOFFMAN, W.C. A design procedure for control/display systems. Human Factors, 1977, 19, 421-436.

DAVIS, G.I. Man-machine interface design using structured analysis and simulation techniques. Proceedings of the IEEE International Conference on Cybernetics and Society, Seattle, WA, October, 1982.

DeGREENE, K.B. Systems analysis techniques. In K.B. DeGreene (Ed.) Systems psychology. New York: McGraw-Hill, 1970.

DRURY, C.G. Human performance in manual process control. In T.B. Sheridan & G. Johannsen (Eds.) Monitoring behaviour and supervisory control. New York: Plenum Press, 1976.

DRURY, C.G. Task analysis methods in industry. Applied Ergonomics, 1983, 14, 19-28.

EASON, K.D. The process of introducing information technology. Behaviour and Information Technology, 1982, 1, 197-213.

FITTS, P.M. (Ed.) Human engineering for an effective air navigation and traffic control system. Washington, DC: National Research Council, 1951. (AD 133 954)

GAEL, S. Problems in evaluating system manning requirements estimates and estimation techniques. Human Factors, 1964, 3, 299-304.

GAGLIARDI, U.O., HUSSEY, R.A., KAPLAN, I.T. and MATTEIS, R.J. Man-computer interactions in idealized tactical problem solving. Darien, CONN: Dunlap and Associates May, 1965. (AD 618 707)

GALER, I.A.R. The role of psychology in the design and evaluation of transport systems. Ergonomics, 1979, 22, 129-133.

GARDNER, K. Human factors in Naval computer-based systems: current possibilities and future trends. Journal of Naval Science, 1982, 8, 253-260.

GARRISON, E.W. Factors in management information system failures. Alexandria, VA: Army Military Personnel Center, December, 1980. (AD A093 102).

GLENN, F.A., ZAKLAD, A.L. and WHERRY, R.J. Human operator simulation in the cognitive domain. Proceedings of the Human Factors Society 26th Annual Meeting, Seattle, October 1982.

GOODBODY, R.L. and MONTELEON, V.J. Navy command, control and communications system design principles and concepts. Vol. 1: C^2 principles and concepts. San Diego, CA: Naval Electronics Laboratory Center, Report No. TD 504, August, 1976. (AD A039 856).

HALLAM, J. and STAMMERS, R.B. The optimum distribution of tasks among operators in a multiman-machine system. Birmingham: University of Aston, Applied Psychology Dept., December, 1979.

REFERENCES (CONT.D)

HANNA, J.F. Information-theoretic techniques for evaluating simulation models. In J.M. Dutton & W.H. Starbuck (Eds.) Computer simulation of human behaviour. New York: John Wiley and Sons, 1971.

HESS, R.A. Aircraft control-display analysis and design using the optimal control model of the human pilot. IEEE Transactions on Systems, Man and Cybernetics, 1981, 11, 465-480.

HOLLNAGEL, E. and WOODS, D.D. Cognitive systems engineering : new wine in new bottles. International Journal of Man-Machine Studies, 1983, 18, 583-600.

HOWELL, W.C. and GETTYS, C.F. Some principles for design of decision systems: a review of the final phase of research on a command-control system simulation. Wright Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratories, Technical Report AMRL-TR-68-158, November, 1968. (AD 684 548).

HOWIE, A. Ergonomics and system design: what price user involvement? In T.J. Triggs (Ed.) Proceedings of the 15th Annual Conference of the Ergonomics Society of Australia and New Zealand, Melbourne, December, 1978.

JOHANNSEN, G. and GOVINDARAJ, T. Optimal control model predictions of system performance and attention allocation and their experimental validation in a display design study. IEEE Transactions on Systems, Man and Cybernetics, 1980, 10, 249-261.

JONES, A.L. and WINGERT, J.W. On estimating the capability of an avionic man-machine system. Proceedings of the IEEE-Ergonomics International Symposium, Cambridge, UK, 1969.

JORDAN, N. Allocations of functions between man and machine in automated systems. Journal of Applied Psychology, 1963, 47, 161-165.

JORGENSEN, C.C. and STRUB, M.H. Analysis of manual threat evaluation and weapons assignment (TEWA) in the AN/TSQ-73 air defence system. Alexandria, VA: U.S. Army Research Institute for the Behavioural and Social Sciences, Report No. ARI-TR-419, October, 1979. (AD A089 612).

KIBLER, A.W., WATSON, S.R. and KELLY, C.W. A prototype aid for evaluating alternative courses of action for tactical engagement. McLean, VA: Decisions and Designs Inc, November, 1978. (AD A064 275).

KIDD, J.S. and VAN COTT, H.P. System and human engineering analyses. In H.P. Van Cott and R.G. Kinkade (Eds.) Human engineering guide to equipment design. Washington, DC: U.S. Government Printing Office, 1972.

KLEINMAN, D.L., BARON, S. and LEVISON, W.H. An optimal control model of human response, part I. Automatica, 1970, 6, 357-369.

KLEINMAN, D.L., PATTIPATI, K.R. and EPHRATH, A.R. Quantifying an internal model of target motion in a manual tracing task. IEEE Transactions on Systems, Man and Cybernetics, 1980, 10, 624-636.

KNOWLES, W.B. Flight controllers for jet transports. Human Factors, 1967, 9, 305-320.

REFERNCES (CONT.D)

KNOWLES, W.B., BURGER, W.J., MITCHELL, M.B., HANIFAN, D.T. and WULFECK, J.W. Models, measures and judgements in system design. Human Factors, 1969, 11, 577-590.

KOLF, F. and OPPELLAND, H.J. Guidelines for the organizational implementation of information systems - concepts and experiences with the PROGI implementation handbook. In N. Bjorn-Andersen (Ed.) The human side of information processing. Amsterdam: North Holland, 1978.

KRAISS, K.F. A display design and evaluation study using task network models. IEEE Transactions on Systems, Man and Cybernetics, 1981, 11, 339-351.

KREIFELDT, J.G. Cockpit displayed traffic information and distributed management in air traffic control. Human Factors, 1980, 22, 671-691.

KUPERMAN, G.G., HANN, R.L. and BERISFORD, K.M. Refinement of a computer simulation model for evaluating DAIS display concepts. Proceedings of the Human Factors Society 21st Annual Meeting, 1977, 305-310.

KURKE, M.I. Operational sequence diagrams in system design. Human Factors, 1961, 3, 66-73.

LANDIS, D., SLIVKA, R.M. and JONES, J.M. Evaluation of large scale visual displays. Griffiss Air Force Base, NY: Rome Air Development Centre Technical Report RADC-TR-67-57, April, 1967. (AD-651 372)

LANE, N.E., STRIEB, M.I., GLENN, F.A. and WHERRY, R.J. The Human Operator Simulator: an overview. In J. Moraal & K.F. Kraiss (Eds.) Manned systems design: methods, equipment and applications. New York: Plenum Press, 1981.

LANE, N.E., STRIEB, M.I. and LEYLAND, W. Modeling the human operator: applications to system cost effectiveness. In Advisory Group for Aerospace Research and Development (AGARD) Conference Proceedings No. 268. Modeling and simulation of avionics systems and command, control and communications systems, 1980. (AD AO82 012/6)

LEAL, A., CHEN, K.K., GARDINER, P.C. and FREEDY, A. Studies and application of adaptive decision aiding in anti-submarine warfare. Woodland Hills, CA: Perceptronics Inc, Report No. PFTR-1035-78-4, April, 1978. (AD AO57 275).

LEVISON, W.H. The optimal control model for the human operator: theory, validation and application. In M.L. Frazier and R.B. Crombie (Eds.) Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics. Edwards Air Force Base, CA: Air Force Flight Test Center Report No. AFTC-TR-82-5, May, 1982. (AD A129 333).

LICKLIDER, J.C.R. Man-computer symboisis. IRE Transactions on Human Factors in Electronics, March, 1960, 4-11.

LINDQUIST, O.H., JONES, A.I. and WINGERT, J.W. An investigation of airborne displays and controls for search and rescue (SAR), volume I: summary. Minneapolis: Honeywell, Inc., Report No. 12609-FRI-vol. 1, July, 1971(a). (AD 728 041).

REFERNCES (CONT.D)

LINDQUIST, O.H., JONES, A.L. and WINGERT, J.W. An investigation of airborne displays and controls for search and rescue (SAR), volume II: SAR requirements and technology survey. Minneapolis: Honeywell, Inc., Report No. 12609-FRI-VOL. II, July, 1971(b). (AD 728 042).

LINDQUIST, O.H., JONES, A.L. and WINGERT, J.W. An investigation of airborne displays and controls for search and rescue (SAR), volume III: avionics analysis and system synthesis. Minneapolis: Honeywell, Inc., Report No. 12609-FRI-vol. III, July, 1971(c). (AD 728 043).

MALTBACK, P.C. Scicon/APU air intercept control skills trainer. London: Scicon Consultancy International Ltd, November, 1980.

MARA, T.D. Human factors in ship control, volume I: analysis of ship operations, operator capabilities and recommended bridge arrangements. Groton, CONN; General Dynamics Corp., January, 1968. (PB 179 356).

McCORMICK, E.J. Human factors in engineering and design. (4th Ed.). New York: McGraw-Hill, 1976.

McCORMICK, E.J. and Sanders, M.S. Human factors in engineering and design. (5th Ed.), New York: McGraw-Hill, 1981.

McCORMICK, E.J. Job analysis: methods and applications. New York: Amacon, 1979.

McRUER, D.T., GRAHAM, D., DRENDEL, E. and REISENER, W. Human pilot dynamics in compensatory systems. Wright Patterson Air Force Base, Ohio: Air Force Flight Dynamics Laboratory, Technical Report AFFDL-TR-65-15, 1965.

MEISTER, D. A pragmatic approach to the prediction of operational performance. In R.E. Blanchard and D.H. Harris (Eds.). Man-machine effectiveness analysis. Los Angeles: Symposium of the Human Factors Society, June, 1967.

MEISTER, D. Human factors: theory and practice. New York: John Wiley and Sons, 1971(a).

MEISTER, D. Comparative analysis of human reliability models. Westlake Village, CA: Bunker Ramo Corp., November, 1971(b). (AD 734 432).

MEISTER, D. Behavioral foundations of system development. New York: John Wiley and Sons, 1976.

MEISTER, D. The role of human factors in system development. Applied Ergonomics, 1982(a), 13, 119-124.

MEISTER, D. Human factors problems and solutions. Applied Ergonomics, 1982(b), 13, 219-223.

MEISTER, D. The present and future of human factors. Applied Ergonomics, 1982(c), 13, 281-287.

MEISTER, D. and FARR, D.E. The utilization of human factors information by designers. Canoga Park, CA: Bunker-Ramo Corp., September, 1966. (AD 642 057).

MEISTER, D. and RABIDEAU, G.F. Human factors evaluation in system development. New York: John Wiley and Sons, 1965.

MEISTER, D. and SULLIVAN, D.J. A further study of the use of human factors information by designers. Canoga Park, CA; Bunker-Ramo Corp., March, 1967. (AD 651 076).

MEISTER, D., SULLIVAN, D.J. and ASKREN, W.B. The impact of manpower requirements and personnel resources data on system design. Wright Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratories Technical Report 68-44, September, 1968.

MEISTER, D., SULLIVAN, D.J., FINLEY, D.L. and ASKREN, W.B. The effect of amount and timing of human resources data on subsystem design. Wright Patterson Air Force Base, Ohio: Air Force Human Resources Laboratory Technical Report AFHRL-TR-69-22, October, 1969. (AD 699 577).

MILLER, D.C., FEEHRER, C.E., MURALIDHARAN, R., PEW, R.W. and BARON, S. Development of human performance models for man-machine system simulation. Cambridge, MA: Bolt, Beranek and Newman, Inc. Report No. 3739, October, 1978. (AD A069 879/5).

MILLER, D.C. and PEW, R.W. Exploiting user involvement in interactive system development. Proceedings of the Human Factors Society 25th Annual Meeting, 1981.

MILLS, R.G. and HATFIELD, S.A. Sequential task performance: task module relationships, reliabilities and times. Human Factors, 1974, 16, 117-128.

MONTGOMERY, C.A., THOMPSON, J.R. and KATTER, R.V. Imagery intelligence (IMINT) production model. Woodland Hills, CA: Operating Systems Inc., Report No. OSI-R79-003, January, 1980. (AD A086 455).

MORGAN, C.T., COOK, J.S., CHAPANIS, A. and LUND, M.W. (Eds.) Human engineering guide to equipment design. New York: McGraw-Hill, 1963.

NADLER, G. The planning and design approach. New York: John Wiley and Sons, 1981.

NICKERSON, R.S. Man-computer interaction: a challenge for human factors research. Ergonomics, 1969, 12, 501-517.

NICKERSON, R.S. and FEEHRER, C.E. Decision making and training: a review of theoretical and empirical studies of decision making and their implications for the training of decision makers. Cambridge, MA: Bolt, Beranek and Newman Inc., August, 1975.

NICKERSON, R.S., MEYER, T.H., MILLER, D.C. and PEW, R.W. User-computer interaction: some problems for human factors research. Cambridge, MA: Bolt, Beranek & Newman Inc, Report No. 4719, September, 1981. (AD A127 089).

REFERNCES (CONT.D)

NICKERSON, R.S. and PEW, R.W.   Person-computer interaction.   In The C^3 system user, vol I: a review of research on human performance as it relates to the design and operation of command, control and communication systems.   Cambridge, MA: Bolt Beranek and Newman Inc, Report No. 3459, February, 1977.   (AD AI26 633).

NTDS FUNCTIONAL ALLOCATION STUDY GROUP.   Final report of the CNO NTDS functional allocation study group.   Washington, DC: March, 1981.   (AD A107 679).

OBERMAYER, R.W.   Simulation, models and games: sources of measurement. _Human Factors_, 1964, _6_, 607-619.

O'REILLY, C.A. and ROBERTS, K.H.   Task group structure communication and effectiveness in three organizations.   _Journal of Applied Psychology_, 1977, _62_, 674-681.

PARKS, D.L. and SPRINGER, W.E.   Human factors engineering analytic process definition and criterion development for CAFES.   Seattle, WA;   Boeing Aerospace Company, January, 1976.   (AD A040 478).

PARRISH, R.N., GATES, J.L. and MUNGER, S.J.   Design guidelines and criteria   for   user/operator   transactions   with   battlefield   automated systems.   Volume 1 - summary.   Fairfax, VA: Synectics Corp, February, 1981(a).   (AD A115 874).

PARRISH, R.N., GATES, J.L. and MUNGER, S.J.   Design guidelines and criteria   for   user/operator   transactions   with   battlefield   automated systems.   Volume 4 - provisional guidelines and criteria.   Fairfax, VA: Synetics Corp, February, 1981(b).   (AD A115 892).

PAYNE, D. and ALTMAN, J.W.   An index of electronic equipment operability: report of development. Pittsbsurgh, PA: American Institute for Research, January, 1962.

PEASE, M.C.   ACS.1:   An experimental automated command support system. _IEEE Transactions on Systems, Man and Cybernetics_, 1978, _8_, 725-735.

PEASE, M.C. and SAGALOWICZ, D.   Study of automated command support systems.   Menlo Park, CA: SRI International, July, 1979.   (AD A072 298).

PEW, R.W. and BARON, S.   Perspectives on human performance modelling.   In Proceedings of the Conferences on Analysis, Design and Evaluation of Man-Machine Systems, Baden-Baden, Germany, September, 1982.

PEW, R.W., BARON, S. FEEHRER, C.E. and MILLER, D.C.   Critical review and analysis   of   performance   models   applicable   to   man-machine   systems evaluation.   Cambridge, MA: Bolt, Beranek and Newman Inc, Report No. 3446.   March, 1977.

PEW, R.W., HOECKER, D.G., MILLER, D.C. and WALKER, J.H.   User interface requirements   for   computer-assisted   delivery   service.   Cambridge, MA: Bolt, Beranek and Newman, Inc, Report No. 4045, June, 1979.

PEW, R.W. MILLER, D.C. and FEEHRER, C.E. Evaluation of proposed control room improvements through analysis of critical operator decisions. Palo Alto, CA: Electric Power Research Institute, Research Project 891, August, 1981.

PEW, R.W., WOODS, W.A., STEVENS, A.L. and WEENE, P. Identifying information requirements for command and control decision making. Cambridge, MA: Bolt, Beranek and Newman, Inc, Report No. 3801, 1978.

PHATAK, A., WEINERT, H. and SEGALL, I. Identification of the optimal control model for the human operator. Wright Patterson Air Force, Ohio: Aerospace Medical Research Laboratories Technical Report AMRL-TR-74-79, May, 1975. (AD-A009 956).

PHELPS, R.H., JOHNSON, E.M. and HALPIN, S.M. A decision support framework for decision aid designers. Proceedings of the IEEE International Conference on Cybernetics and Society, Denver, October, 1979.

PICKREL, E.W. and McDONALD, T.A. Quantification of human performance in large complex systems. Human Factors, 1964, 6, 647-662.

POTEMPA, K.W., LINTZ, L.M. and LUCKEW, R.S. Impact of avionic design characteristics on technical training requirements and job performance. Human Factors, 1975, 17, 13-24.

PRICE, H.E., FIORELLO, M., LOWRY, J.C., SMITH, M.G. and KIDD, J.S. Department of defence and service requirements for human factors R & D in the military system acquisition process. Falls Church, VA: Biotechnology Inc, July, 1980. (AD A107 891).

PRITSKER, A.B., WORTMAN, D.B., SEUM, C.S., CHUBB, G.P. and SEIFERT, D.J., SAINT: volume 1. Systems analysis of an integrated network of tasks. Wright Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratories Technical Report AMRL-TR-73-126, April, 1974. (AD-A014 843).

PROCTOR, C.L. and KHALID, T.M. A quantitative approach to performance evaluation of man-machine systems having a stochastic environment. International Journal of Man-Machine Studies, 1971, 3, 127-140.

PYLYSHYN, Z.W. Computational models and empirical constraints. Behavioural and Brain Sciences, 1978, 1, 93-127.

RAMSEY, H.R. and ATWOOD, M.E. Human factors in computer systems: a review of the literature. Englewood, CO: Science Applications Inc., Report No. SAI-79-111-DEN, September, 1979. (AD A075 679).

RAN SCHOOL OF TRAINING TECHNOLOGY. Royal Australian Navy training system manual. Volume 2 - 1 - analysis for training. Royal Australia Navy School of Training Technology, HMAS CERBERUS, Victoria, Australia: May, 1978.

RASMUSSEN, J. The human as a systems component. In H.T. Smith and T.R.G. Green (Eds.) Human interaction with computers. London: Academic Press, 1980.

REFERNCES (CONT.D)

REBANE, G.J., WALSH, D.H. and MOSES, F.L.   Tactical on-line maneuver
model.   Proceedings of the IEEE International Conference on Cybernetics
and Society, Denver, October, 1979.

REGULINSKI, T.L.   Quantification of human performance reliability:
research method rationale. In J.P. Jenkins (Ed.) Proceedings of the U.S.
Navy Human Reliability Workshop, Washington, DC, July, 1970.   (AD 722
689).

ROSSMEISSL, P.G. and DOHME, J.A.   Using rating scales to determine
aptitude requirements of Army systems. In Proceedings of the 24th Annual
Conference of the Military Testing Association.   Brooks Air Force Base,
Texas: Air Force Human Resources Laboratory Report No. AFHRL-TP-83-16,
March, 1983. (AD A126 554).

ROSS, D.T. and SCHOMAN, K.E.   Structured analysis for requirements
definition. IEEE Transactions on Software Engineering, 1977, 3, 6-15.

ROUSE, S.H. and ROUSE, W.B.   Evaluation of an onboard computer-based
information system for aircraft.   Proceedings of the IEEE International
Conference on Cybernetics and Society, Atlanta, GA, October, 1981.

ROUSE, W.B.   Human-computer interaction in multi-task situations.   IEEE
Transactions on Systems, Man and Cybernetics, 1977, 7, 384-392.

ROUSE, W.B.   Systems engineering models of human-machine interaction.
New York: North Holland, 1980.

SAMET, M.G. and DAVIS, K.B.   Computer-based supervisory system for
managing information flow in C^2 systems: pacing model. Woodland Hills,
CA: Perceptronics Inc, Technical Report PTR-1033-77-3, March, 1977. (AD
A040 359/2).

SAMET, M.G., WELTMAN, G. and DAVIS, K.B.   Application of adaptive models
to information selection in C^2 systems.   Woodland Hills,   CA:
Perceptronics Inc., Technical Report PTR-1033-76-12, December, 1976. (AD
A038 523/7).

SAWYER, C.R., FIORELLO, M., KIDD, J.S. and PRICE, H.E.   Measuring and
enhancing the contribution of human factors in military system
development: case studies of the application of impact assessment
methodologies.   Falls Church, VA: Biotechnology Inc., July, 1981.   (AD
A122 031).

SENDERS, J.W.   The human operator as a monitor and controller of multi-
degree of freedom systems.   IEEE Transactions on Human Factrors in
Electronics, 1964, 5, 2-5.

SHERIDAN, T.B. and FERRELL, W.R.   Man-machine systems: information,
control and decision models of human performance.   Cambridge, MA: MIT
Press, 1974.

SIBERT, J.   Top-down methodology for human factors research.   In C.M.
Mitchell, P.M. Van Balen and K.L. Meo (Eds.) Human Factors Cosnsiderations
in System Design, NASA Conference Publication 2246, January, 1983. (N83-
18238 to N83-18250).

REFERNCES (CONT.D)

SIEGEL, A.I. and FEDERMAN, P.J. Validation of the DEI technique for large-scale display evaluation. Rome, NY: Rome Air Development Centre, Technical Report RADC-TR-67-134, 1967.

SIEGEL, A.I. and FEDERMAN, P.J. Communications content training as an ingredient in effective team performance. Ergonomics, 1973, 16, 403-416.

SIEGEL, A.I., FISCHL, M.A. and MACPHERSON, D. The Analytic Profile System (APS) for evaluating visual displays. Human Factors, 1975, 17, 278-288.

SIEGEL, A.I., LEAHY, W.R. and LAMB, J. Human reliability in a combined sonar system: operator reliability for a selected AN/SQS-26 and AN/SQR-10 scenario. Wayne, PA: Applied Psychological Services, September, 1976(a).

SIEGEL, A.I., HEAHY, W.R. and LAMB, J. Human reliability in a combined sonar system: human reliability for the AN/SQS-26, LAMPS and AN/SQR-19 system. Wayne, PA: Applied Psychological Services, December, 1976(b).

SIEGEL, A.I., LEAHY, W.R. and WIESEN, J.P. Applications of human performance reliability evaluation concepts and demonstration guidelines. Wayne, PA: Applied Psychological Services, March, 1977. (AD A037 632).

SIEGEL, A.I., LEAHY, W.R. and WOLF, J.J. A computer model for simulation of message processing in military exercise control and evaluation systems. Wayne, PA: Applied Psychological Services, October, 1977. (AD A045 832).

SIEGEL, A.I., LEAHY, W.R and WOLF, J.J. Human performance tradeoff curves for use in the design of Navy systems. Wayne, PA: Applied Psychological Services, April, 1978. (AD A053 332).

SIEGEL, A.I., MADDEN, E.G. and PFEIFFER, M.G. Evaluations of operational decision aids, 3 - evaluative approaches and methods. Wayne, PA: Applied Psychological Services, Inc, August, 1980. (AD-A103 444).

SIEGEL, A.I., MADDEN, E.G. and WOLF, J.J. Application of computer simulation techniques in military exercise control system development. Wayne, PA: Applied Psychological Services, May, 1981. (AD A126 248).

SIEGEL, A.I., MIEHLE, W. and FEDERMAN, P. The DEI technique for evaluating equipment systems from the information transfer point of view. Human Factors, 1964, 3, 279-286.

SIEGEL, A.I. and WOLF, J.J. A technique for evaluating man-machine system designs. Human Factors, 1961, 3, 18-28.

SIEGEL, A.I. and WOLF, J.J. Man-machine simulation models. New York: John Wiley and Sons, 1969..

SIEGEL, A.I. and WOLF, J.J. Digital behavioural simulation: state-of-the-art and implications. Wayne, PA: Applied Psychological Services, June, 1981. (AD A128 641).

SIEGEL, A.I., WOLF, J.J. and WILLIAMS, A.R. A model for predicting integrated man-machine system reliability. II Model validation. Wayne, PA: Applied Psychological Services Inc, March, 1976. (AD A023 886).

REFERNCES (CONT.D)

SILVERN, L.C. Training: men-man and man-machine communications. In K.B.De Greene (Ed.) *Systems psychology*. New York: McGraw-Hill, 1970.

SINGLETON, W.T. *Man-machine systems*. London: Penguin Education, 1974.

SMITH, R.L., WESTLAND, R.A. and CRAWFORD, B.M. The status of maintainability models: a critical review. *Human Factors*, 1970, 12, 271-283.

SMITH, S.L. Requirements definition and design guidelines for man-machine interface in C^3 system acquisition. Bedford, MA: Mitre Corp, Report No. M80-10, June, 1980. (AD A087 258).

SMITH, S.L. Man-machine interface (MMI) requirements definition and design guidelines; a progress report. US Air Force ESD-TR-81-113-MTR 8134, 1981.

STEEB, R., ARRTOF, M., CROOKS, W.H. and WELTMAN, G. Man/machine interaction in adaptive computer-aided control: final report. Woodland Hills, CA: Perceptronics Inc., Report No. PATR-1008-75-12/31, December, 1975. (AD A021 323).

STEEB, R., CHEN, K. and FREEDY, A. Adaptive estimation of information values in continuous decision making and control of remotely piloted vehicles. Woodland Hills, CA: Perceptronics Inc., Report No. PATR-1037-77-8, August, 1977. (AD A050 414).

STEEB, R., CHU, Y., CLARK, C., ALPEROVITCH, Y. and FREEDY, A. Adaptive estimation of information values in continuous decision making and control of advanced aircraft. Woodland Hills, CA: Perceptronics Inc., Report No. PATR-1037-79-6, JUNE, 1979. (AD A077 917).

STEWART, T.F.M. Displays and the software interface. *Applied Ergonomics*, 1976, 7, 137-146.

SWAIN, A.D. Some problems in the measurement of human performance in man-machine systems. *Human Factors*, 1964, 6, 687-700.

TAINSH, M.A. On man-computer dialogues with alpha-numeric status displays for Naval command systems. *Ergonomics*, 1982, 25, 683-703.

TAINSH, M.A. Job process charts and man-computer interaction within naval command systems. Teddington, Middlesex: Admiralty Marine Technology Establishment, July, 1983 (see also *Ergonomics*, 1985, 28, 555-565).

THORNDYKE, P.W. and WEINER, M.G. Improving training and performance of navy teams: a design for a research program. Santa Monica, CA : Rand Corp., Report No. R-2607-ONR, July, 1980. (AD A089 092).

TONG, R.M. ARBEL, A., CIOFFI, S.O., KELLEY, C.M., PAYNE, J.R. and TSE, E. Option generation techniques for command and control. Mountain View, CA: Advanced Information and Decision Systems, January, 1983. (AD A127 715).

REFERNCES (CONT.D)

TOPMILLER, D.A. A factor analytic approach to human engineering analysis and prediction of system maintainability. Wright Patterson Air Force Base, OH: Aerospace Medical Research Laboratories, Report No. AMRL-TR-64-115, 1964.

TOPMILLER, D.A. Methods: past approaches, current trends and future requirements. In J. Moraal and K.F. Kraiss (Eds.) Manned systems design: methods, equipment and applications. New York: Plenum Press, 1981.

TRIGGS, T.J. The role of protocol analysis in understanding human cognitive processes. Cambridge, MA: Bolt, Beranek and Newman, Inc., Report No. 2616, July, 1973.

TUCKER, J.H. Implementation of decision support systems. Proceedings of the IEEE International Conference on Cybernetics and Society, Georgia, October, 1980.

TULLIS, T.S. An evaluation of alphanumeric, graphic and color information displays. Human Factors, 1981, 23, 541-550.

TVERSKY, A. and KAHNEMAN, D. Judgement under uncertainty: heuristics and biases. Science, 1974, 185, 1124-1131.

U.S. GENERAL ACCOUNTING OFFICE. Guidelines for assessing whether human factors were considered in the weapon system acquisition process. Washington, DC: December, 1981.

Van COTT, H.P. and KINKADE, R.G. (Eds.) Human engineering guide to equipment design. Washington, DC: U.S. Government Printing Office, 1972.

Van HORN, R.L. Validation of simulation results. Management Science, 1971, 17(5), 247-258.

VAUGHAN, W.S., WHITTENBURG, J.A. and GILLETTE, P.R. Basic concepts for future command control system planning. Alexandria, VA: Whittenburg Vaughan Associates, Report No.W/V-RR-66/1-Cd, December, 1966. (AD 816 149).

VITTAL, J., SELFRIDGE, O. and BOBROW, R. Using AI techniques for threat display and projection, including tactical deception indication. Cambridge, MA: Bolt, Beranek and Newman, Inc., Report No. 4789, September, 1981. (AD A106 801).

WALDEN, R.S. and ROUSE, W.B. A queuing model of pilot decision-making in a multi-task flight management situation. IEEE Transactions on Systems, Man and Cybernetics, 1978, 8, 867-875.

WESSON, R.B., HAYES-ROTH, F., BURGE, J.W., STASZ, C. and SUNSHINE, C.A. Network structures for distributed situation assessment. IEEE Transactions on Systems, Man and Cybernetics, 1981, 11, 5-23.

WHERRY, R.J. The Human Operator Simulator-HOS. In T.B. Sheridan and G. Johannsen (Eds.) Monitoring behaviour and supervisory control. New York: Plenum Press, 1976.

REFERNCES (CONT.D)

WIENER, E.L. and CURRY, R.E.   Flight deck automation promises and problems. Ergonomics, 1980, 23, 995-1011.

WOHL, J.G. Force management decision requirements for air force tactical command and control. IEEE Transactions on Systems, Man and Cybernetics, 1981, 11, 618-639.

WOHL, J.G.   Information automation and the Apollo program: a retrospective. IEEE Transactions on Systems, Man and Cybernetics, 1982, 12, 469-478.

WOHL, J.G., ENTIN, E.E., ALEXANDRIDIS, M.G. and ETERNO, J.S.   Toward a unified approach to combat system analysis. Burlington, MA: Alphatech, Inc., January, 1983. (AD-A124 570).

WOODSON, W.E.   Human factors design handbook.   New York: McGraw-Hill, 1981.

WORTMAN, D.B., SEIFERT, D.J. and DUKERT, S.D.   SAINT simulation of a remotely piloted vehicle/drone control facility - model development and analysis.   Wright Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratories, Technical Report AMRL-TR-75-118, June, 1976. (AD A031 085).

ZACHARY, W.W.   Decision aids for Naval air ASW.   Willow Grove, PA: Analytics, Technical Report 1366-A, March, 1980. (AD A085 134).

APPENDIX A

User Involvement in Systems Design

(a) Introduction

It is virtually a truism to say that a comprehensive programme of human factors input during system specification will require some form of involvement of the users of the proposed system. Human Factors Engineering, by definition, should be concerned with the performance and comfort of individuals within the operational system. This focus on the human factor demands that both the performance attributes and the desires of the proposed users should be communicated to the design team. In practice, this communication may be achieved in a number of ways. One common method is that the human factors specialist performs a liaison function between users and design engineers. In some situations, users have not been consulted directly. In the worst case, the design team itself may generate the user requirements. This decision process may be based on such factors as experience with past systems, intuition or 'common knowledge'.

It may be appreciated that, while user involvement is often regarded as axiomatic (Nadler, 1981), the nature and the extent of this involvement vary considerably. A central theme of this appendix will be that user involvement should be given a relatively high priority at an early stage in the design process in order that complex human-machine systems, such as C^3 systems, may be designed effectively. The first section will investigate more explicitly the benefits of user involvement in system specification. A brief survey and evaluation of the methods of promoting this involvement will be made. Types of user involvement will be investigated, as will be the problems associated with communicating user requirements to the design team.

(b) Benefits of user involvement

From an historical perspective, a consideration for the requirements of the users of man-machine systems may be seen as arising from the increasing complexity of those systems. For example, Singleton (1974) distinguishes the hardware-centred approach to design from the more comprehensive systems approach to design, that necessarily includes a consideration of the operator. He claims that the systems approach has been stimulated by experiences in which improvements in technology have failed to increase system effectiveness. Further, in the view of Bjorn-Andersen (1978) amongst others, attempts to improve the performance of individual operators through the methods of traditional ergonomics may not be enough. He believes that the introducers of complex technologies must also address any social issues that are likely to result. For example, in the field of management information services, it may be that employee dissatisfaction translates rather directly into costs of absenteeism and staff turnover.

The relevance of this latter point to the issue of user involvement in the design of C^2 systems may not be obvious. However, social factors do have an influence upon C^2 system performance, an example being the possible preference of operators for shared VDU's in contrast to individual units (see Cohen & Turney (1972) for further details).

This emphasis on social concerns at least suggests that the era in which designers could generate user requirements through conventional wisdom is becoming increasingly remote.

User involvement in system design is often cited as an ideal, and the subsequent benefits are often presumed rather than specified. Yet one large benefit that cannot be ignored is that early consultation with the proposed users may identify problems which may have otherwise only become apparent in the operational system (Miller and Pew, 1981; Eason, 1982). As was discussed in Section 1.1, early human factors input to the system design process may help to prevent costly re-designs at a later stage. Insofar as user involvement should be included in programs of human factors input, the latter activity implies the former. A related advantage of user involvement is that early quantification of important parameters of the system may be achieved (Miller and Pew, 1981). This last paper also addressed the social aspects of information systems, in that it was claimed that user 'sponsorship' of the final system may be important.

(c) Types of user involvement

As mentioned previously, the nature of user involvement in the system development process varies considerably. In this report, a number of workers have emphasised the importance of the degree of active participation of the potential users, e.g., Howie (1978), Eason (1982). That is, design procedures which merely pay 'lip-service' to consulting system users are likely to be unsatisfactory. As an example of such an approach, Howie (1978) refers to the 'hostage' method in which a representative user is placed on the design team but is given neither education nor influence.

The example raises a second issue, namely, how many of the potential users of a system should be consulted during design. If one wishes to define 'users' as anyone who will have direct contact with the new system, it is possible that the amount of effort expended to consult all users during design may be enormous. Clearly, representative users must be chosen (see Section 6.9). This implies that the consultation of some user groups is more important than that of others, and also that different user groups should be consulted at different stages of the design process (Nadler, 1981). An important user group often overlooked during the design process is that responsible for maintenance of the operational system (Brooks, Grouse, Jeffrey and Lawrence, 1982).

Possibly the most crucial feature of active user participation is that comments should be sought regarding the desirability of a number of alternative designs (Miller and Pew, 1981; Eason, 1982). In fact, one may extend this participation further by allowing the users to become responsible for the generation of some of the options available to them (Howie, 1978). Obviously, this approach requires that users possess above-average design and evaluation skills. In the field of information systems, at least, Eason's (1982) concept of evolutionary design specifically allows for user education through the progressive implementation of parts of the system.

(d) Methods of promoting user involvement

The techniques of promoting user involvement in system design constitute one of the general methods of implementing human factors input. As the methods and techniques for achieving this latter goal are the primary concern of this whole report, the review and evaluation at this stage will be brief, particularly where these methods overlap.

A number of conceptually useful distinctions may be made between the various methods of promoting user involvement. Firstly, in C^2 system design it should be noted that not only must hardware be developed, but also suitable procedures and software have to be devised in order that the system functions effectively. Given a system specification in which the hardware details are largely predetermined (for reasons which take no account of human factors, e.g. hardware cost/availability), the role of the user will be limited to assisting in 'soft' design. The disadvantages of hardware pre-determination have been discussed previously; (in Section 1.1) namely, fitting users to hardware designs constrains the development process. In the field of information systems, however, it may well be that the greatest benefit of involving users lies in their assistance with solving the software and associated procedural problems.

A second useful distinction concerns the manner in which one elicits information from users regarding their requirements in the proposed system. Basically, one has the option of canvassing user opinion directly through the use of questionnaires, group interviews, debriefings, etc., or alternatively, of obtaining some form of objective performance measure through user interaction with a previous or prototypal system. The latter method naturally revolves around a contrived systems test, and utilises such techniques as task and protocol analysis, and work sampling.

A useful evaluation of these methods has been made by Ramsey and Atwood (1979). Briefly, those methods that rely on the canvassing of opinion suffer from the subjective nature of the data obtained. Whilst users may be expert at performing their jobs, this is no guarantee that they will be as competent at analysing and verbalising their performance. In fact, oft-repeated procedures may become so automatic that they cease to be regarded as significant. These methodological problems may be overcome partially through structured interviews, but the data are still subjective. By contrast, the objective methods may be more reliable, but suffer other problems. When observing users as they interact with a system, there is the danger that current (inappropriate) practices may be enforced. Secondly, it is difficult to observe covert behaviour, such as occurs in many cognitive tasks. The ideal solution is possibly a combination of subjective and objective methods, such as a performance test followed by (or including) debriefing or protocol analysis.

With regard to the desirability of active user involvement in systems design, it would appear that eliciting direct user opinion fulfills this end most suitably. On the other hand, even the techniques based on questionnaires and interviews have been criticised for inviting 'outside' control of the system development process (Miller and Pew, 1981). Active user analysis of present or prototypal systems has been advocated both because more constructive information is said to be gained (Miller and Pew, 1981) and because of general concerns for industrial democracy in management information systems (Kolf and Oppelland, 1978).

Once again, this method presumes a technically sophisticated user.

In practice, the distinction between direct and indirect methods of involving users is unlikely to produce radically different system designs. Whether users are questioned directly or subjected to performance assessment, a large degree of interpretation of the resulting data occurs. Expert guidance of the system development process is inevitable, given non-expert users. Possibly, a goal which is both desirable and practical is that the methods that are used to elicit user information are not unduly influenced by the preconceptions of the designers. The best means of achieving this end is probably the use of a number of different methods simultaneously, cf., Pew, Hoecker, Miller and Walker (1979); Nadler (1981). The use of these multiple methods appears to be a feature of the systems approach to design (Singleton, 1974), in which the design process itself is becoming subjected to increased rigour.

(e) Problems of user involvement

It has been pointed out several times throughout this appendix has been that active user participation in systems design requires a resonable level of user sophistication. User naivete is probably the greatest barrier to more active involvement. Eason's (1982) concept of an evolutionary design procedure attempts to reduce this last problem. In the meanwhile, it is likely that the best compromise is to employ a human factors engineer as a liaison between users and designers.

User ignorance of design manifests itself in a number of ways. If one questions users a priori about their requirements, it is likely that the resulting answers will be so broad as to make the process of translating these requirements into specific hardwares and/or procedures difficult. Users tend a priori to suggest designs which are either minor variations on a familiar system, or which are grossly unrealistic (Miller and Pew, 1981). User exposure to some form of prototype or simulation appears to yield the most constructive data (Miller and Pew, 1981). In this context, it is likely that the ability of users to evaluate alternative designs is much greater than their ability to generate these designs themselves. This disability contributes to user neglect in the design process, as the process of asking for criticism of a prototype constrains the possible replies.

A second class of problems encountered in user involvement may broadly be termed 'professional'. That is, resistance from both management and design engineers may occur towards efforts for user participation. User involvement may result in time delays, which have to be justified (Miller and Pew, 1981). Secondly, the greater the number of interests represented on the design team, the more the potential for conflict (Howie, 1978; Eason, 1982).

APPENDIX B

Studies of the Design Process

Investigations of the manner in which design engineers work are useful for gaining an understanding of methods which can promote human factors input during system development. In this context, the timing of the presentation of human factors data to the designers appears to be crucial. Typically, the major conceptual aspects of designs are formulated at an early stage of the system development process (Meister, Sullivan and Askren, 1968). Hardware configurations tend to become fixed at a relatively early stage, and later work revolves around the embellishment of this configuration with finer detail. This result provides support for what would otherwise be an oft-repeated article of faith: that human factors should be a consideration in the design process from the start. As discussed previously, the effectiveness of human factors input is limited in a situation where the hardware configuration is fixed.

Secondly, given that timely design-related human factors data are available, the style of presentation of such data has been investigated in relation to its utilisation by design engineers (Meister, Sullivan, Finley and Askren, 1969). In particular, the research was designed to investigate the utility of incrementally-presented human factors data on the final design (of the propellant transfer and pressurisation subsystem of the Titan III project). The incremental nature of data presentation is a common feature of system development cycles, i.e. the designers' requests for human factors guidelines typically become more detailed as development proceeds. However, it was shown (Meister et al, 1969) that simultaneous presentation of all human factors related data yielded superior designs. Further investigation uncovered some reasons for this finding.

Firstly, design proceeded so rapidly that incremental human factors inputs tended to lag behind, and were subsequently ignored. Secondly, an attitude survey revealed that designers had difficulty in conceptualizing the impact of human factors except at a molecular level, i.e. in the design of controls and displays. For example, manning requirements tended to be regarded as outgrowths from equipment design, rather than as design constraints themselves. Simultaneous presentation of all human resources data (including personnel numbers and training levels) was seen as overcoming some of these conceptual difficulties.

Studies of design engineer behaviour (Meister and Farr, 1966; Meister and Sullivan, 1967) have tended to confirm what has long been suspected: that designers make less than optimum use of human factors data (at least when laying out a hypothetical control panel). As a consequence, there may be heavy reliance on experience rather than analysis during the design process, and a dominant, perhaps inappropriate concern for hardware details. These findings alone do little to suggest how professional indifference to human factors may be overcome. However, further work did uncover some reasons.

Neglect of human factors may be said to result from a combination of designer attitude and inappropriate presentation of human factors data

(Meister et al, 1968).   As regards designer attitude, it appears once
again that engineers have difficulty in assessing the impact of human
factors except at a molecular 'knobs and dials' level. The result is that
human factors guidelines may only be heeded at a relatively late stage of
design. Hardware is seen as defining the constraints which will be placed
on humans in the system (if this issue is considered at all), rather than
the design being subject to human requirements.

On the other hand, there is also evidence that human factors
guidelines must be framed in a certain manner in order for designers to be
responsive.   In particular, constraint-related information appears to be
required.   Engineers need to be impressed with the probable consequences
of ignoring a human-based guideline, preferably in a quantitative
fashion.   In this context, Meister and Sullivan (1967) criticised the
then-current U.S. Military Specifications for being overly general and
qualitative.   Secondly, it may be that graphical presentation of
information has a greater impact on subsequent design than the more
popular textual presentation (Meister and Farr, 1966).

Generally, designers have difficulty translating the more
abstract human factors requirements into specific designs. For example,
spatial constraints are easily applied to interface design, whilst
relatively great difficulty may be experienced when trading off concepts
of personnel numbers and training levels. One solution, according to
Meister and Farr (1966), may be to provide designers with better means of
analysing conceptual systems. The goals of this report are thus closely
aligned with that philosophy.

END

DATE
FILMED

7 8 8