THE UNIVERSITY OF HULL

Human Skill Maintenance in Complex Work Environments: Applications to Extended Spaceflight

being a Thesis submitted for the degree of

Doctor of Philosophy

in the University of Hull

by

Juergen Sauer M.Sc.

March 1997

Acknowledgements

I would like to thank Prof Bob Hockey for his supervision and invaluable support throughout the completion of this PhD. Furthermore, I wish to thank my second supervisor, Dr David Wastell, for writing the experimental software and for providing helpful advice during the course of the research.

This research work was completed part-time while I was employed as a research assistant in the Psychology Department at the University of Hull. I am grateful to the European Space Agency, in particular, Mr A Bichi from ESA-LTPO, for funding parts of the research reported in this thesis. The work was also supported indirectly by various other contracts from the University of Hull, without which, it would not have been possible to continue the work for the PhD.

My thanks also go to the many people from the Psychology Department at Hull University who contributed in various ways to the successful completion of this work. In particular, I would like to thank Nick Shryane, Iain Coyne, Patricia Taylor-Shipley, Andrew Geake, Jonathon Downing and Kevin Thomas for reading sections of the thesis and providing helpful comments. Furthermore, thanks are due to Linda Randerson and the staff from the workshop, Ray Wallis, Trevor Cook, Peter Fountain, Pam Knott, Sandra Readhead and Bob Richmond, for their help during the completion of the research.

Finally, I am grateful to my wife Larissa and to my parents for their support and encouragement.

TABLE OF CONTENTS

Chapter 1: Introduction 1

Chapter 2: Automation of Complex Work Environments 19

Chapter 3: Methodology 36

Chapter 4: Simulation Task 54

Chapter 6: Knowledge and Training 100

Summary of thesis

This thesis examines human performance under sub-optimal working conditions during work with complex and highly-automated process control systems. The operational context focuses on applications in extended spaceflight but the generic approach allows for generalisations beyond this target work environment.

The methodological approach is based on the use of a computerised multiple-task environment to carry out generic simulations of real work environments (micro-worlds) with a high level of ecological validity. For that purpose, a PC-based task environment was developed to simulate the operation of a life support system in a spacecraft. This task environment has been used in lab-based experiments with trained participants from the student population and with real space crews during large-scale mission simulations. A series of six experiments was carried out (3 lab and 3 field studies) to investigate the impact of different configurations of sub-optimal working conditions and unfavourable operator states, using the following independent variables: sleep deprivation, dialogue control, social isolation and confinement, training, noise, extended lay-off period and different types of system faults (corresponding to variations in workload). The task environment comprised up to five tasks, allowing for the observation of differential effects of the independent variables on different levels of cognitive activity. Dependent variables included primary task performance, secondary task performance, system control behaviour, information sampling behaviour, and subjective state measures.

The findings suggested that primary performance was rarely affected, whereas certain secondary task measures and, notably, information sampling strategies appeared to be good indicators of changes in demand under the unfavourable conditions. The isolation and confinement experiments revealed no serious breakdown of performance among the crew but some indications of strain were observed. The use of two different training approaches displayed a very complex picture, with no method showing clear superiority over the other concerning performance, though there were differences in knowledge structure and system management behaviour. An important implication of the experimental work is that a broad methodological approach is needed in order to investigate the complex adjustment patterns displayed by individuals during the management of task demands under unfavourable conditions.

Chapter 1:

Introduction

1.0 Summary

This chapter outlines the remit of the thesis with its focus on human factors aspects of extended spaceflight. As many other work environments, spaceflight has enjoyed an increasing level of automation, which has given rise to a number of problems associated with highly-automated systems. In addition, because of a number of particular environmental features, notably those of isolation and confinement, a higher risk of performance decrements is associated with spaceflight than with similar terrestrial work environments. This risk is expected to rise significantly with increasing mission length.

The result of a literature-based task analysis suggested that a large part of crew time is spent on process control-type activities, supporting the assumption that there are strong similarities between spaceflight and industrial process control. A review of the literature on human factors issues in space travel revealed a clear shortage of research in this area. This may partly be compensated for by employing suitable analogous work environments. An evaluation of a number of candidate environments indicated that nuclear submarines, industrial process control and underwater habitats provided the closest match to space regarding the technical environment. Antarctic expeditions (and to some extent mountaineering expeditions) have often been cited in the literature as suitable analogous environments. While they are similar to space with regard to isolation and confinement features, they clearly lack the complexity of the technical systems, thus being only of limited use.

1.1 Plan of the thesis

The thesis investigates human performance and behaviour during interaction with highly-automated and complex process control systems, such as those experienced in extended spaceflight. The investigations are carried out under various sub-optimal working conditions and operator states, which are relevant performance shaping factors in this work environment. The operational context is based on an application of process control, where highly-automated life support systems are used to maintain suitable atmospheric conditions in a space vessel.

Two trends in manned space travel are giving rise to problems, which have not been sufficiently addressed in space research. First, space missions have increased in length, from the Apollo moonlanding mission of 10 days in 1969 to the stay of the Russian cosmonaut Valerij Polyakov on MIR station for 438 days in 1994/5. However, future missions are likely to exceed this period by far, with a Mars mission, for example, expected to take at least two years. The second trend is the increasing proliferation and complexity of automated systems in the space environment, partly driven by the need to make most efficient use of resources (oxygen, fuel, food etc.) through the employment of complex recycling processes to reduce the payload.

Increased levels of automation in many terrestrial work environments (industrial process control, aviation, air traffic control etc.) have already given rise to a number of problems (e.g. inappropriate trust in reliable functioning of system, loss of manual control skills) associated with working with highly-automated and complex systems. A large number of scientific publications have addressed these issues in a generic manner (e.g. Reason 1990, Wickens 1992) as well as in the context of particular work environments such as nuclear power plants (Woods, O'Brien and Hanas 1987), aviation (Wiener and Nagel 1988) and air traffic control (Wise, Hopkin and Smith 1991). Similar problems are therefore likely to be faced in space, in particular, with the increasing duration of missions.

Therefore, the work embodied in this thesis attempts to contribute to closing the gap between human factors requirements of extended spaceflight and the research carried out in that field so far. More precisely, it attempts to achieve the following goals. First, it aims to achieve a better understanding of how human performance during work with complex systems is affected by performance shaping factors relevant to extended spaceflight. Second, it tries to achieve a better understanding of the strategies humans use to cope with excessive work demands. Third, it aims to provide some answers to the many problems faced for work design during extended space travel. As part of the work, a software package has been developed to provide a generic simulation of some aspects of the cognitive work environment of space. The simulation task was used to investigate a number of issues, which have been associated with performance impairment in the literature. The overall goal was to look at behaviour in a complex and dynamic decision-making environment with often conflicting goals rather than isolating any particular aspects of the task environment to test specific hypotheses.

Chapter 1 provides a general introduction to the problems of human skill maintenance in the context of extended spaceflight. Based on an analysis of the space work environment, a number of issues are identified, which are considered to be performance shaping factors. Those issues selected for empirical investigation are reviewed in the respective experimental chapters while problems associated with automation and their implications for work design are discussed in Chapter 2. Being aware of the issues related to automation is fundamental to understanding the problems of other performance shaping factors in space as well as in many other work environments. Chapter 3 describes the methodological framework of the thesis and explains the rationale for selecting the performance shaping factors examined. Chapter 4 gives an account of how the methodological framework guided the design of the task environment and provides a detailed technical description of the different versions of the simulation task used in the empirical work. Chapters 5-7 present the six experiments carried out for the thesis and include a review of the literature, addressing the main independent research variables examined in the respective experiments. Based on the empirical findings of this thesis and results from the research literature, Chapter 8 discusses the implications, first for the theories and models upon which this research was based, and second for practitioners who are engaged in the design of the space work environment or some analogue work environments.

1.2 General problem domain

The fact that missions have become longer has added a qualitatively new dimension to spaceflight, with a range of issues now becoming predominant, which played only a minor role in short space missions. These are issues associated with the psychological effects of long-term isolation and confinement. The same orbital working and living environment is much less likely to pose any problems during short missions than during long flights since humans are able to tolerate even very unfavourable conditions for a limited period of time (Hunt 1987).

The list of potential stressors in spaceflight is long and each of them may represent a serious threat to mission success as a single agent, with additional interaction effects to be expected from the presence of several stressors. Micro-gravity causes a number of physiological changes (e.g. decreased bone size, congestion, space sickness), which affect well-being and may, in turn, impinge on performance (Bluth and Helppie 1987). The working and living environment is deprived of the usual stimulants found in equivalent terrestrial environments. The crew members are socially isolated and confined to very small spaces, which they are unable to leave for a long time. The number of direct social contacts is not only limited but also unchanging (the work colleagues are the same people one also spends one's leisure time with). The reduced level of stimulation also applies to the internal habitat, of which the interior decor remains virtually constant. Space travellers are exposed to a constant level of noise (up to 70 dB on some missions), which is caused by the continuous operation of the life support systems (Connors, Harrisson and Akins 1985). Furthermore, they have to live under very difficult nutritional and hygienic conditions (e.g. no showers).

This brief description has demonstrated the potential risk of a serious performance impairment, or even a complete breakdown, as a result of prolonged exposure to these stressors over the course of the mission. Unlike in most terrestrial work environments, crew members are not easily replaceable if a serious performance breakdown occurs. The possibilities of alleviating stress are more restricted than in earth-based environments since a physical withdrawal from work is impossible and many stress relieving activities cannot be carried out in a space capsule.

1.3 Description of space work environment

A literature-based task analysis of the space work environment has been conducted to improve the understanding of the cognitive tasks the crew carries out during the course of a mission. The results have also represented the basis for the design of the simulation task used in the experiments. Obviously, missions differ greatly with regard to a number of characteristics such as duration, purpose and task allocation. Therefore, an attempt has been made to give a more general description of the tasks and to evaluate their overall significance for crew survival.

Crew work activities were summarised into seven categories (discussed in turn below). The analysis was based on a number of sources (Kass 1990; Gushin 1994; Connors et al

1985; Bluth and Helppie 1987), which provided information about the task structure of different missions completed by ESA, NASA and the Soviet Space Agency.

Management/Planning. This category refers to the short- and long-term planning of activities. The nominal daily schedule of a mission day often has to be modified according to situational requirements (e.g. repetition of an unsuccessfully completed experiment). Certain events may also necessitate the change of long-term plans. For example, the illness of a crew member may require the complete abandonment of some activities and the distribution of others to the rest of the crew. Management and planning activities are always carried out in co-ordination with ground control, who usually have a major input into the scheduling of activities. Daily reports also have to be written to provide detailed information about the state of the crew and about progress with the schedule.

System control. The activities in this category refer to the management of the different technical systems in the spacecraft, of which most are critical for crew survival. A spacecraft is usually equipped with the following systems: Life Support and Environmental Control, Electrical Power, Thermal Control, Data Management, Communication, and Warning (Kass 1990). The crew has to ensure a safe functioning of all the systems, which is a key task for mission success. A part of system control activities is the completion of maintenance tasks, which may be subdivided into preventive and corrective maintenance. Consumables (e.g. fluids) have to be replaced from time to time, the system has to be cleaned regularly, and handover procedures to the next 'shift' have to be completed. Furthermore, the training of emergency procedures has to be carried out at certain intervals to ensure the maintenance of crucial skills. Overall system control activities represent the group of tasks, which take up the largest proportion (about one third) of working time (Gushin 1994).

Navigation of spacecraft. Although the spacecraft is largely flying automatically, during certain mission phases, the pilot takes control over the vessel and carries out different manoeuvring operations. The model of spaceflight being similar to ‘flying’ a plane may be obsolete since manual navigational activities are generally infrequent and only to be performed during manoeuvres such as docking procedures, orbital corrections and landing. Nevertheless, monitoring is still required to ensure that the spacecraft remains on course, which makes the task similar to the situation in very modern aircraft. In these aircraft strategic flight management is more frequent than tactical manoeuvring (Billings 1991). This manifests itself in pilots often 'programming' their way out of a problem rather than adopting manual control. Although overall time requirements for the completion of these activities are rather low compared to total mission time, the activity is regarded as critical for mission success.

Payload operations. In this context the concept of payload refers to the total amount of scientific investigations that are to be carried out during a mission. Payload operations mainly involve the preparation, completion and evaluation of experiments. Furthermore, the experimental facilities have to be regularly inspected, serviced and repaired if necessary (Kass 1990). It is sometimes required to integrate new facilities or reconfigure the equipment already installed. Some of the experimental work requires the availability of longer periods of uninterrupted crew time since the full attention of the astronaut is needed for successful completion. Although some of the experiments may involve a considerable level of risk, for example, when handling toxic materials, payload operations are usually not critical for mission safety. Nevertheless they take up a considerable proportion of working time and often represent the actual purpose of the space mission. Over recent years one has also witnessed increased automation of payload operations, requiring little crew member input, to increase the number of scientific experiments that can be completed during a mission (Kass 1990; Wichman 1990).

EVA (Extravehicular activity). Although the majority of activities is intravehicular, a number of tasks have to be completed outside the space capsule (e.g. docking manoeuvres, visit of satellites). EVAs are difficult to carry out since a space-suit has to be worn, which impedes the astronaut's movements. Furthermore, life risk is elevated because the activity takes place outside the protecting space capsule and a number of life-threatening events are possible (e.g. being hit by a meteorite, space-suit malfunction or depressurisation, disconnection from tether cable,(Bluth and Helppie 1987). An EVA can last for up to 8 hours, which includes all pre- and post- preparatory activities. Since EVAs consume an enormous amount of physical energy, they are not completed very frequently (not more than twice a month on most Soviet missions).

Communication. In addition to the communication requirements between crew members, frequent communication via different channels (audio-, video-, computer-based) with ground control is required to coordinate activities and keep all parties informed. Selecting the best means of communication for the kind of information to be transmitted can be a crucial factor for enhancing communication effectiveness. In comparison with most terrestrial environments, communication in space is associated with greater difficulties. The principal problem is the degradation of voice communication through high levels of ambient noise and artificial atmospheres (e.g. low air pressure or unusual gas mixtures, Connors et al 1985). This is a considerable drawback since voice communication is often the preferred form of communication in situations (e.g. repair work, payload operations) where the use of both hands is required. Research has even been conducted into the possibility of using space systems that can be voice-activated by the crew (Kass 1990).

Training. On-board training represents a very important activity to maintain essential skills. There are a number of skills that might be at risk from degradation, in particular those, which have not been practised during long periods of time (e.g. skills required during re-entry into terrestrial atmosphere). Although advancements in computer technology have considerably facilitated the possibilities of providing effective training programmes during in-flight, careful planning is still required to ensure that the skills most at risk are allocated the appropriate training time. However, in future long-duration missions one may also schedule the acquisition of some (non-safety critical) skills for in-flight training rather than for pre-mission training (Connors et al 1985). Not only would this reduce pre-mission training time but it may also relieve some of the boredom typically associated with certain mission phases (e.g. cruise phase).

This brief review illustrates that spaceflight comprises a wide range of very varied activities, with a strong emphasis on process control-related operations including system control, fault identification, maintenance and repair. Space activities also seem to differ with regard to safety criticality. While some operations are of very high importance for crew survival, others, such as payload operations, tend to be less critical. Furthermore, the intervals at which activities are carried vary considerably. Certain tasks are characterised by very long time intervals between two consecutive task executions or are even completed only once during a mission. This has implications for an evaluation of training needs since low frequency tasks are subject to a higher risk of skill decrement than tasks that are completed regularly and at short time intervals.

1.4 Review of research on human performance in space

Spaceflight, as a research field, is an interdisciplinary domain, requiring the collaboration of many scientific disciplines in order to minimise the enormous risks associated with this undertaking. Consequently, a large amount of research has been carried out in different disciplines, which contributed to the remarkable achievements in this area. However, in comparison with engineering and many of the pure sciences, focused psychological research has been very scarce.

In the beginning of the space era, the focus was on short-term adaptation problems of the crew to the unusual work and living environment (Connors et al 1985). With increasing mission length the focus has changed to problems associated with isolation and confinement but it did not address the issue in the context of a highly-automated work environment. Put differently, within psychology the emphasis has been on the social and physical aspects of the space environment while neglecting the technical work environment. The lack of human factors research is probably a reflection of the general difficulties of this discipline to gain recognition as an important contributor to system design (Sanders and McCormick 1992). Nevertheless, a small amount of research has been carried out on performance aspects of the technical environment of spaceflight. This is reviewed here.

1.4.1 Performance decrements and errors

The overall occurrence of serious accidents in spaceflight has been quite low, compared to other work environments such as aviation (Nagel 1987). This has been attributed to the fact that crew members had to undergo rigorous and well-designed selection and training procedures. Furthermore, their level of motivation has been extremely high, which also helped to prevent errors. However, this mainly refers to data from short missions and one may therefore not conclude that during extended spaceflight human error will also remain at the same level. Another factor may have contributed to the low number of reported human errors in spaceflight. Space travellers were found to avoid, if possible, informing ground control about their errors in order not to endanger their participation in future missions (Gushin 1992). Nevertheless some performance decrements have been observed. They may be classified into two types according to their primary cause: Adaptation problems to the space environment and effects of an extended stay in space.

Although the first type of decrement is somewhat better researched than the second, because of the experience gained in short flights by the American and Soviet space agencies, only a few relevant studies are available. This has partly been attributed to the fact that simulators were primarily used for training purposes rather than for carefully designed performance assessment studies (Connors et al 1985). The results from US space missions suggested that the crew generally had difficulties in reaching pre-flight performance levels during the first mission days (Connors et al 1985). This was mainly due to adaptation difficulties to micro-gravity and the physiological problems associated with it (e.g. space sickness). This resulted in more time needed to complete tasks in space than during the last terrestrial mission simulation. Similarly, Soviet studies suggest that during the first nine days of the mission a general deterioration of mental efficiency occurs (Gushin 1991). There was also some evidence that previous spaceflight experience was a crucial factor in determining performance at the beginning of the flight (Bluth and Helppie 1987). Cosmonauts flying for the first time suffered more strongly from a general reduction in information processing speed than their colleagues with previous flight experience. One of the very few systematic attempts to study human factors issues in space was an experiment by Manzey, Lorenz, Schiewe, Finell and Thiele (1995), which examined an astronaut's performance on a dual task (tracking and memory search) during an 8-day space mission. The findings indicated that dual-task performance as well as single-task tracking was impaired during the flight compared to pre- and post-mission baseline levels, whereas single-task memory search performance was unaffected. This underlines the considerable effect of micro-gravity on sensorimotor skills.

Knowledge is even more limited for the second type of decrement, since there has been little research of any kind on performance during extended spaceflight. Although a considerable number of isolation studies in spacecraft simulators (lasting between 7 and 365 days) have been conducted over the last 30 years in the former Soviet Union

(Gushin, Kholin and Ivanovsky 1993), these have focussed largely on biomedical and psychological problems but have not addressed performance as a primary issue. Three recent isolation studies (ISEMSI, EXEMSI and HUBES[1]) carried out by the European Space Agency (ESA) have attempted to redress the balance by making psychological issues a major focus of the programme (Bonting 1993, 1996; Værnes, Baranov, Demin and Stepanov 1995). A review of individual experiments conducted during the studies is presented in Chapter 7.

1.4.2 Workload and Fatigue

Workload varies a great deal in spaceflight, especially on long missions, with extended periods of low workload during the cruise phase and periods of very intense work activities during landing and docking operations. Both types of deviation from normal workload levels are associated with negative effects but the effects may be qualitatively different.

In order to avoid the problems associated with underload (such as drifting attention, missing signals, lethargic performance), Soviet cosmonauts were kept busy during those periods (Connors et al 1985; Bluth and Helppie 1987). Furthermore, during periods of low workload, crew members were found to report increased feelings of loneliness (Bluth and Helppie 1987).

Excessive work demands are generally associated with stress responses of various kinds, which may result in performance decrements, changes in work strategies and fatigue (Hockey 1993). There was considerable evidence in the Soviet space programme - based on real spaceflight and simulation studies - that fatigue was a problem during the mission (Gushin et al 1993). This required, for example, a temporary change of the work/rest schedule on some Soviet missions, where allocated sleeping time was increased to 12 hours (Bluth and Helppie 1987).

A number of interviews carried out with Soviet and European space crews also suggested that increasing fatigue during the mission has been a common problem (Kass 1990). Similarly, two isolation studies (ISEMSI and EXEMSI) carried out by ESA in a hyperbaric chamber gave some indication of the occurrence of fatigue. In a four-week study (ISEMSI), a marked effect of cognitive fatigue was observed in one crew member during the final week of isolation (Hockey and Wiethoff 1993). The second isolation study (EXEMSI), which lasted for 8 weeks, also indicated an increase in fatigue (albeit small in magnitude) over the isolation period for most crew members (Hockey and Sauer 1996).

1.5 Analysis of analogue work environments

The review of the human factors literature on space has shown that there is a paucity of research in this area. This was the principal reason to widen the outlook to various other work environments, which can be argued to be similar to space in some ways. The approach of using analogue environments to advance knowledge in a new research field has been widely used in spaceflight (Harrison, Clearwater and McKay 1990; Connors et al 1985). A considerable number of analogues have been suggested in the literature as being suitable for application to space. The most frequently discussed analogue environments are: Aviation, Air Traffic Control, Industrial Process Control, Antarctic Expeditions, Underwater Habitats, Nuclear Submarines and Mountaineering Expeditions. Although other environments (such as prisons, monasteries, speleological missions, oil drilling in deserts or in the sea) have also been discussed as representing suitable space analogues (Lavitola, Tomatis, Loria and Pinotti 1990), they were not included in the analysis. This was because they were judged to be too dissimilar to space travel since their commonalities with space travel were based solely on a certain degree of social isolation (and confinement in some cases) found in these analogues.

In order to obtain a better understanding of the degree of similarity between space and the analogue work environments, a literature-based evaluation was carried out. Three criteria were used for this, representing different aspects of the operational environment. *Technical environment* includes factors such as workload, degree of automation and the type of tasks (monitoring, problem-solving, etc.) to be carried out. *Ambient environment* includes issues such as noise levels, habitability and confinement features. *Social environment* refers to factors such as the number of colleagues at work, the diversity of their professional background, the degree of social isolation and the extent of team work required.

All seven analogous environments have been assessed on these three dimensions in terms of their similarity to space. The degree of similarity was indicated by a three-star rating (* = low similarity, ** = medium similarity, *** = high similarity). The analysis suggested that nuclear submarines, industrial process control and underwater habitats had the closest match to space with regard to the technical environment. Aviation, ATC and Antarctic expeditions were found to be less similar, whereas mountaineering expeditions were quite dissimilar to space on that dimension. However, Antarctic and mountaineering expeditions (and also underwater habitats and nuclear submarines) generally appeared to be good analogues regarding the ambient and social environment. Table 1.1 summarises the results of this evaluation.

Before each work environment is discussed in more detail, some qualifications need to be made concerning the significance and purpose of the evaluation. First, the analysis is rather crude since it was developed on the basis of three raters, making a literature-based assessment of each work environment. Second, an additional difficulty encountered was the variability of the characteristics within each work environment. For example, spaceflight differs according to the nature of the mission (interplanetary flight vs extended stay on space station) and so do most other analogue environments (civilian vs military aviation, short vs long-haul flights, etc.). Third, the analysis did not attach differential weights to the three dimensions, which may be inappropriate.

Analogue	Dimension of Work Environment		
	Technical	Ambient	Social
Aviation	**	**	*
Air traffic control	**	*	*
Industrial process control	***	*	*
Antarctic expeditions	**	***	***
Underwater habitats	***	***	***
Nuclear submarines	***	***	**
Mountaineering expeditions	*	**	**

Table 1. 1: Evaluation of Analogue Work Environments (Degree of similarity to space: *** = high, ** = medium, * = low)

The purpose of the evaluation was not to provide a detailed and comprehensive comparison of space travel with the analogue work environments. Instead, it was aimed to obtain a broad understanding of the strengths and weaknesses of each candidate environment as an analogue to space. Some confirmation of the overall appropriateness of the rating was obtained during the presentation of the findings to an audience of international experts belonging to the psychological advisory board of the European Space Agency (Wastell and Sauer 1995). Some modifications of the ratings were made as a result of this discussion.

Aviation. The operational environment of aircraft crews has undergone considerable changes in recent years, moving towards increasing levels of automation, which resulted in drastic changes of the pilot's role, from one of active flying to that of a supervisor of largely self-flying machines (Wiener 1988, Billings 1991). Whereas the physical demands of flying the aircraft have continually decreased over the years, the cognitive demands have risen drastically with more complex systems being introduced and a larger number of instruments to be monitored (Sexton 1988). Compared to the other work environments discussed here, the job of aircraft crews requires fast decision-making in response to a rapidly changing external environment. Furthermore, the level of 'forgiveness' of mistakes is comparatively low since some errors may not be easily correctable in the time available. Although aviation is highly automated like the space environment, the nature of the work is only moderately similar since ‘flying’ represents only a minor aspect of the crew member’s work. With only short-term isolation and confinement found in aviation, this feature of the environment shows little similarity to space.

Air Traffic Control. The job of an air traffic controller requires a sustained high level of performance with considerable decision-making pressure (Harss, Lichtenfeld, Kastner and Goodrich 1991). Sustained concentration is necessary since tasks often need to be carried out over extended periods with few breaks. The work of an air traffic controller is very complex since information has to be drawn from various sources (e.g. traffic conditions, weather conditions, communication with pilots and controllers from other sections) to arrive at an appropriate decision concerning the routing of an aircraft. The level of automation is only moderate with automatic devices introduced only more recently (see, for example, Lemoine and Debernard 1996). Job control is rather high, allowing the air traffic controller to adopt workload reducing strategies (e.g. the temporary 'parking' of aircraft in designated holding areas) to avoid any serious performance degradation (Spérandio 1971). Overall, similarity to space is only moderate.

Industrial Process Control. Very high levels of automation have been reached in a number of industrial process control applications such as chemical and nuclear power plants. As a result of higher automation and increasing system complexity (with an enormous number of system components and cross-coupled variables), operators have been largely removed from the control loop, being left with monitoring tasks to ensure that the automatic system is working properly (Woods, O'Brien and Hanes 1987). The variables operating in process control systems are rather slow with long time constants (Wickens 1992). Therefore, action taken by the operator usually receives feedback from the system with a time lag. During preventive and corrective maintenance activities, a great deal of communication is required between the operators and the maintenance engineers since they may be physically far removed from each other (operators remain in control room, engineers are on site) with little knowledge of each others activities (De Keyser 1986). Concerning the technical aspects of work, a considerable degree of similarity exists with space travel, whereas little similarity was found on the other dimensions.

Antarctic expeditions. In the context of work environments, this refers to scientific over-wintering camps on the South Pole. The period of isolation and confinement is the most extended of all analogues analysed in this review. It lasts between 7 to 9 months and typically corresponds to the duration of the Antarctic winter. However, the degree of confinement is less severe with the polar base providing considerably more space than underwater habitats or space stations (Lavitola et al 1990). During over-wintering, some risk is involved due to harsh climatic conditions and non-rescuability of the crew if the necessity arises (e.g. serious illness). Overall, Antarctic expeditions provide a good analogue as for the ambient and social environment. However, the level of automation and complexity of the technical systems in Antarctic over-wintering camps is moderate at best (though, increasing), which makes it quite dissimilar to space.

Underwater habitats. With regard to the social and ambient work environment, there appeared to be a considerable degree of similarity to spaceflight, with a typical underwater station (such as Sealab) having a similar layout like a space station (Welham

1994). Underwater habitats also require a closed-loop life support system to maintain appropriate atmospheric conditions (Kass 1990). Although the technological systems used in underwater habitats may be slightly less complex than those in space vessels, a considerable degree of similarity exists between the two. Missions in underwater stations typically last for up to 15 days (e.g. Sealab programme) with the exception of the TEKTITE project, where the crew was living in the hyperbaric habitat for 60 days (Lavitola et al 1990). Diving missions into the surroundings of the habitat are an important part of the work and essentially correspond to EVAs (extravehicular activities) in space missions. Overall, underwater habitats represent a very good analogue to the space environment since they show a high degree of similarity on all three dimensions.

Nuclear Submarines. Modern nuclear submarines are equipped with highly complex technological systems, which support four major functions: Navigation, life support, nuclear reactor and weapon guidance systems (Welham 1994; Encyclopaedia Americana 1986). Life support systems (e.g. maintenance of temperature and humidity, air purification, extraction of O_2 from sea water) and the nuclear reactor are highly-automated systems, which show a strong similarity to industrial process control and spaceflight activities. Navigation of a submarine resembles to some extent flying an aircraft with similar types of rudders available, which manoeuvre the submarine into different horizontal and vertical directions. Periods of submersion can be of considerable length (e.g. global submersive circumnavigation by US submarine Triton in 84 days), during which the crew is exposed to a substantial degree of confinement. Social isolation is less severe than in the other isolated environments since social contacts are more varied with a crew of up to 170 members. Like in space vessels, the crew is exposed to constant noise. Even in peace times, a not inconsiderable life threat is associated with each mission due to factors such as accidental pressurisation, power loss at great depths, radiation and contamination through gaseous substances (Weybrew 1990). Like underwater habitats, nuclear submarines score high on all dimensions, which also makes them good analogue environments. One may consider underwater habitats as the analogue to space stations and nuclear submarines as the analogue to interplanetary flights[2].

Mountaineering Expeditions. This operational environment is characterised by low tech equipment in comparison with all other analogous environment, which makes it very dissimilar to space regarding the technical aspects. However, it has been cited in the literature as a suitable analogue because of the isolation features and the danger associated with it. Expeditionary ascents involve a considerable degree of danger because of the very difficult conditions of the ambient environment (low temperatures, strong winds, avalanches etc.) representing a permanent risk of injury, disease or death (Encyclopaedia Americana 1986). Similarly as in spaceflight, expedition members may suffer from mountain sickness in high altitudes during the first few mission days. However, generally its usefulness as an analogue appears to be rather limited.

The results of the comparative analysis indicated the close technical similarity of industrial process control, underwater habitats and nuclear submarines to spaceflight. This provides a starting point for a search of the research literature in these analogues for material relevant to extended space travel. This may help us to improve our understanding of the problems of humans interacting with complex technical systems during exposure to difficult ambient and social work conditions. However, a review of the research literature revealed that only industrial process control provides ample material on these issues while only a modest amount of information is available on the two other environments. In both cases, the emphasis has been on physiological and medical problems associated with stays in great depths rather than the interaction with a complex technical work environment (Weybrew 1990). If the study included a psychological perspective, however, it largely focussed on psychiatric episodes of crew members not being able to cope with the environmental demands (Weybrew and Noddin 1979; Kanas 1985). Overall, little is published about the human-machine interaction in these environments[3]. Two reasons may account for this. First, this may be symptomatic for the problems of human factors research to gain recognition as an important contributor to system design (see Chapter 2). Second, because of the predominantly military nature of the work (this applies particularly to nuclear submarines), some of the research findings may not be available to the general public. This leaves process control as the main information source. However, all other analogues that showed sufficient similarity to certain aspects of the technical environment are also made use of where appropriate. The implications of complex and highly-automated systems for effective human-machine system performance are explored in the next chapter.

Endnotes:

---

[1] The goal of all three studies was broadly to increase our knowledge of the effects of extended human involvement in a space station environment. All studies were conducted in a hyperbaric chamber (ISEMSI & EXEMSI) or a terrestrial mock-up of a space station (HUBES) and investigated medical, biological, physiological and psychological issues. ISEMSI (Isolation Study for the European Manned Space Infrastructure) took place in Bergen (Norway) in 1990 and lasted for 4 weeks with a crew of six. EXEMSI (Experimental Campaign for the European Manned Space Infrastructure) was conducted in Cologne (Germany) in 1992 and lasted for 8 weeks with a crew of four. HUBES (Human Behaviour in Extended Spaceflight) took place in Moscow (Russia) in 1994-95 and lasted for ca. 19 weeks with a crew of three.

[2] The propulsion system for a Mars mission may consist of a nuclear reactor since this is expected to lead to a significant reduction in traveling time, leaving more time for the exploration of the planet (Murray 1993).

[3] Weybrew (1990) provides a brief summary of the human factors issues researched since the beginning of the nuclear submarine era without providing further references. The summary is not of sufficient detail to be of use in the present case but it indicates that some human factors research did take place, focussing on noise and display aspects of the system.

Chapter 2:

Automation of Complex Work Environments

2.0 Summary

There has been a marked increase in automation in many complex work environments. This also applies to space travel, where many systems (such as navigation, life support and sometimes even payload operations) have become automated. Although in most cases the benefits of automation outweigh its drawbacks by far, considerable attention needs to be paid to the effects on human operators and their role in the human-machine-system. This chapter discusses those issues in automation and system design that are considered critical to improving the effectiveness of human-machine-systems.

The level of control the operator enjoys over the system and the degree of transparency of the system's activities will strongly influence the operator's situation awareness and the effectiveness with which manual control can be applied. There is a considerable risk of skill loss if manual control skills are not regularly practised, which is often the case in highly automated and highly reliable systems. The prevalence of low-frequency tasks in spaceflight exacerbates this problem. The frequency with which manual system interventions are carried out is also influenced by the operator's trust in the reliability of the system and the operator's confidence. The availability of appropriate operator support systems may considerably enhance overall human-machine-system performance. The proliferation of automatic devices affects the kind of human error occurring and increases the magnitude of potential disasters. Automation tends to increase variation in workload levels, with workload being only moderate under normal operational conditions but rising to extremes under emergencies. The chapter concludes with the system design implications of the differences found between space and the analogue environments.

2.1 Introduction

Owing to the widespread use of automated devices and systems in space (in navigation, life support, payload operations), it was felt that a separate chapter on automation was needed to discuss the implications for work design and operator performance. Many of the pertinent issues encountered in the analogue work environments are also relevant to space, though, because of the differences between them, a different emphasis is required.

Rapid technological progress over recent years has led to significant changes in many work environments. Increased automation is probably the most notable of these developments, resulting in considerable changes in the type of tasks carried out by the human operator. Behind these changes were a number of driving forces, which exerted continuous pressure on the status quo (Morgan 1986). There was primarily a need for increased productivity, as a result of the underlying logic of the capitalistic system (which also applied to non-capitalistic state economies to allow them to compete with the capitalist world). Automation allowed for the achievement of more standardised production processes and, most importantly, for a more standardised output, resulting in better quality control (Child 1984). It also overcame some of the limitations of the human operator, with regard to physical, physiological and information processing capabilities.

These technological changes had a number of undoubtedly desirable effects in that many repetitious routine tasks as well as hazardous tasks have been taken over by the machine. On the other hand, this has also caused serious problems at an individual as well as organisational level since these new technical systems have often not worked as efficiently as expected (Wickens 1992; Wall and Davids 1992). Work design was frequently driven by technological (and also organisational) requirements and did often neglect the perspective of the human operator who had to cope with inadequately designed jobs (Spacapan and Oskamp 1990). Although it is often thought that technological developments drive work organisation and system design, the idea of 'technological determinism' appears to be a myth (Wall 1987; Child 1984). In fact, a great deal of discretion exists with regard to the allocation of functions within a system. Using design guidelines and workload assessment techniques, functions may be allocated in various ways (variantly or invariantly to human or machine) to optimise human-machine-system-performance.

Generally, the issue of automation is relevant to space as much as to most analogue work environments. Automation is high and widely used in the key areas of crew activities, such as navigation, life support, payload management facilities (Kass 1990; Connors et al 1985) and is expected to become even more widespread in the future (European Space Agency 1994). However, there are a number of factors referring to the human-machine-system and to the external environment, which highlight important differences such as operator qualification and three-way allocation of function. All these points are discussed in this chapter and, in particular, how well the knowledge gained from research into analogue work environments can be applied to spaceflight.

2.2 Advanced process control as a model work environment

Within the group of analogue work environments available, process control was considered to be particularly suitable. First, it has close similarities to a number of space systems and represents the type of work on which the experimental simulation task is modelled. Second, a large body of research is available in the area of process control. This is in contrast to the other highly suitable analogue work environments identified in Chapter 1, nuclear submarines and underwater habitats, where very little research is available. In order to better understand the processes and principles on which the experimental simulation task is built, it was felt appropriate to give a brief description of typical activities in process control.

The areas of application for process control cover a wide range of industrial activities; e.g. chemical production, power plants, distribution of electrical power, water treatment plants and steel production. Although the areas of application are very diverse, they share at least four general characteristics (Wickens 1992). First, the process variables controlled have long time constants and therefore respond only very slowly to operator input. Second, the variables controlled represent an analogue and continuous process and not a discrete and symbolic one, as found in advanced manufacturing. Third, a large number of variables have to be controlled in the process, of which many are cross-coupled. This means that changes in the state of one variable will have an impact on others, and the underlying dynamic process is often difficult to understand because of the resulting high complexity of the system. Fourth, many areas of process control application represent high risk industries, such as nuclear and chemical plants. This places an additional burden on the operator since an accident could potentially result in a disaster of unlimited scope (e.g. Three Mile Island, Bhopal).

The job of a process control operator involves a variety of tasks (see Daniellou 1986 for a description). The kinds of task carried out depend on a number of factors, such as the operational environment (e.g. nuclear power plant, electricity distribution, chemical industry), the size of the system (work distribution between staff) and the level of automation (the extent to which manual control is required). An attempt is made to provide a generic description of the key tasks carried out by a typical process control operator. One may make a distinction between control modes (automatic or manual) the system operates in. In the automatic mode, the operator is mainly carrying out monitoring activities, together with record-keeping and preventive maintenance activities - often supervised by the control room operator and completed by on-site maintenance staff. The operator's activity pattern is different in manual control mode, where one may distinguish between scheduled manual control (SMC) and unscheduled manual control (UMC). SMC usually involves activities such as start-up and shut-down procedures whereas UMC refers to unplanned events such as recovering the system from a disturbance. Manual control activity is required in both cases but UMC often involves the additional identification and location of system faults. These diagnostic activities sometimes require real-time problem-solving, for example, when a novel fault has occurred in the system. The manual control activities may also involve real-time problem-solving if no procedure-based control strategies are available to deal with the disturbance. SMC, on the other hand, typically involves the execution of well-established procedures according to an instruction manual. However, in practice it may not always be possible to follow these procedures (e.g. new equipment often does not behave the same way as well used machines) so that knowledge-based system control may be required. Overall, the job of a process control operator requires a considerable amount of knowledge and skill to cope with nominal as well as off-nominal situations.

2.3 System design issues

System design is an activity that is typically characterised by a cycle, comprising a number of feedback loops that modify the system concept and its fabrication at various stages in the process (Wickens 1992). A number of important issues in system design are now considered with a particular emphasis on the requirements for spaceflight.

2.3.1 Level of control (out-of-the-loop)

An important issue in the design of automated systems is the level of control allocated to the operator. It refers to the degree of decision-making power in the process, which may range from full manual control (no system assistance) to full automation (no human intervention) with different degrees of control in between (Endsley and Kiris 1995). Since full automation was striven for by many system designers, mainly for reasons of eliminating human error, many operators have found themselves in the position of a *supervisory controller* (Sheridan 1987, Moray 1986). This means that human intervention is not necessary if the system satisfactorily performs the specified tasks but it is still required if things do go wrong. Because of the opaqueness of the system (resulting from the lack of feedback provided), the operator is often unaware of the current system state, which makes it difficult to select an appropriate strategy to deal effectively with an emergency (Wickens 1992; Wiener 1987, 1988). Not being able to take over control of the system in time has already been attributed to a number of accidents in aviation and process control (Nagel 1988). In addition, it may result in monotony and boredom with a consequent loss of task involvement by the operator. It is certainly inappropriate to claim, 'the more control for the operator, the better' as a panacea since there are certain tasks that cannot be controlled manually. To give an example, in the case of an automatic control system failure, astronauts would not have been capable of riding the manned booster rocket in the Apollo programme (Wickens 1992); thus there was no need to give the crew control over that function. Instead, one should strive to provide the appropriate level of control, based on the preferences of the user and the requirements for human-machine-system-performance. Given the high level of qualification and skill of space crews, one would expect them to have a greater need to be in control of system operations than the average process control operator. A more detailed discussion of dialogue control and interface issues is presented in Chapter 5.

2.3.2 Situation awareness

Being out-of-the-loop has also been associated with a loss of *situation awareness* (Endsley and Kiris 1995). As a global concept being applicable to many work environments, situation awareness (SA) is defined as 'the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future' (Endsley 1988). This implies that SA may be described at three levels: (1) All relevant elements in the environments are identified (e.g. parameters in a process control task) and (2) the meaning of these is understood in the light of current operator goals (e.g. appropriateness of current parameter levels to achieve production goals). (3) The highest level of SA refers to the prediction of future states of these elements (e.g. prediction of fluid levels in 30 minutes if leak continues). SA corresponds essentially to one dimension of the mental model concept, namely the *mental model of the current system state* (Wirstad 1988). As situation awareness, this emphasises the importance of making correct interpretations of the current state (level 2) and the ability of making accurate predictions of the future system state (level 3).

Although SA is currently a very popular concept in the literature (e.g. a special issue of the Journal 'Human Factors', Vol. 37 (1) 1995, was dedicated to this subject), it does not seem to be fundamentally different from the ideas discussed in the mental model literature. Nevertheless, the idea of SA is very important in understanding human behaviour in this context and the meaning of SA is broader than its mental model equivalent since it refers to the current situation in the environment (external) rather than the current system state (internal). The connotation of situation awareness, as opposed to the mental model of the current system state, has made the concept probably more appealing for applications to rapidly changing environments such as aviation.

2.3.3 Monitoring and vigilance

The new role of a supervisory controller places more emphasis on monitoring activities. However, humans generally do not perform well on tasks of this kind. It has been demonstrated in vigilance research that attention is difficult to sustain for long periods of time (Parasuraman 1986). Signal detection rates often decline as a function of time, sometimes only 30 minutes into the task. Vigilance decrements are also increased by reduced salience of targets, time or location uncertainty of targets and an increased rate of non-targets. Maintaining performance on vigilance tasks is also associated with a high consumption of mental resources (Deaton and Parasuraman 1988). In response, alarms have been used to alert the operator of malfunctions in the system. However, alarms are rarely completely reliable so that the operator is still required to carry out some monitoring. Furthermore, because of the strong interconnectedness of modern systems, any deviation from a target state will result in a large number of alarms going off within a very short time (alarm inflation syndrome), which are uninterpretable for the operator without appropriate historical information.

Monitoring performance has also become more complicated because of the increased proliferation of components in highly-automated systems (Wickens 1992). Operators do not only have to monitor the possible failure of the initial system (e.g. heater in a spacecraft) but also the automated device that controls the heater and the supervisory monitor that checks the faultfree operation of both systems. The supervisory monitor could also fail in two ways (false alarm or miss), which complicates the situation even further for the operator.

As in many terrestrial automated systems, the monitoring load is considerable in spaceflight. However, in contrast to most analogue environments, many different, unrelated systems are to be monitored (e.g. life support, navigation, payload facilities), which requires a number of independent mental models. However, unlike in nuclear submarines, where one finds a similar arrangement of independent systems, current requirements for spaceflight do not allow for the same degree of specialisation.

2.3.4 Loss of skill

The role of a supervisory controller has a further important implication. It prevents the operator from regularly practising his skills. This imposed passive role is associated with the long-term effect of skill deterioration because of the lack of involvement with direct control (Billings 1991). As a result, pilots have been found to routinely make use of the manual mode in order to maintain their flight skills (Wiener 1988). In process control, operators have even been found to carry out unauthorised manual overrides of the automatic system (Hockey and Maule 1995). In some work environments, such as aviation, the problem of long-term skill deterioration is being addressed by using aircraft simulators to practise manual flight operations. However, not all work environments have these excellent facilities at their disposal so that loss of skill as a consequence of automation is a serious problem (especially in safety-critical work environments).

Loss of skill gains a further dimension in space travel because skills only relevant to certain mission phases (e.g. landing) may not be practised during long periods of time (e.g. cruise phase). This extended lay-off period leads to an increased risk of performance decrements, which represents a greater problem in space than in many other work environments. However, research has been very limited in the area of long-term skill retention (Hagman and Rose 1983, Annett 1989). It is necessary to identify the skills that are particularly at risk, then design an optimal in-flight training programme.

2.3.5 Trust and confidence

The use of manual control is also influenced by trust and confidence in one's own abilities. Trust into a system is developed as a function of the reliability of that system (Muir 1988). Reduced system reliability increases the propensity of operators to control the process manually. Trust is considered to have strong all-or-none characteristics; once it is lost it takes a long time to recover. The second factor, the operator's self-confidence, also plays an important role in determining the likelihood of operator intervention (Lee and Moray 1994). Research has shown that operators often tended to over-estimate their control skills (Kleinmuntz 1990, Moray and Lee 1990), resulting in more interventions than were judged necessary. One is therefore faced with the following dilemma. Whereas a certain amount of trust in the automatic controller is to be regarded as positive, too much trust has been identified as potentially very risky since this may lead to a false sense of security (Wiener 1988). An example of this is the use of the Ground Proximity Warning System (GPWS), initially introduced in aircraft following a series of crashes based on altimeter errors to give a warning signal when the plane comes too close to the ground. However, the crew sometimes over-relied on this aid so that they failed to carry out routine checks on the position of the aircraft in reference to the ground. Similarly, the introduction of a collision warning system in air traffic control has led to a number of incidents, where air traffic controllers left the correction of the flight direction of aircraft until the very last moment although the aircraft were on a direct collision course for some time (Wickens 1992). This risky behaviour was attributed to an over-reliance on the automatic system, which was expected to issue a warning signal before a collision could occur. The issues surrounding trust, mistrust and overtrust are very complex, resulting in great difficulties in finding the appropriate level. An even finer balance is to be struck in space travel since the level of danger associated with this kind of work requires that more emphasis is given to establishing trust during crew training, stressing the protective properties of the space vessel (Mundell 1993).

2.3.6 Allocation of functions

Similar as in work teams, where decisions need to be made about which task is to be completed by whom, the system designer is required to decide which functions are to be carried out by the machine and which by the human operator. Early attempts to provide answers to the issue of allocation of functions produced long lists of functions assigned to either the human (e.g. inductive reasoning, innovation) or the machine (e.g. repetitive activities) after having compared the two with regard to skills and abilities on each task (e.g. Fitts 1951). These lists were adapted as the capabilities of machines improved, allowing them to take over more and more functions from the human operator (e.g. Bekey 1970). The idea of compiling lists of functions and assigning them to either the human or the machine was criticised because the two were not really comparable (Jordan 1963). Human and machine performance cannot be assessed in the same way since they operate fundamentally differently. Consequently, Jordan argued that they should be regarded as complementary rather than being subject to a flawed comparison. However, despite the obvious sense in his arguments, these lists (or tables of relative merit) have been widely used ever since for pragmatic reasons. This was because an ineffective tool at hand was regarded as being better than having no tool at all (Kantowitz and Sorkin 1987).

Another argument was put forward, which questioned the idea of a rigid assignment of tasks. Since workload levels (e.g. increased task demands) as well as operator state (e.g. fatigue) keep changing during a work session, a more flexible approach would bring benefits to the operator in that he/she would be able to reduce workload during periods of high demand (by passing on the job to the machine) and to take on additional tasks during slack times to avoid the negative impact of underload. This approach has been named *dynamic allocation* (Kantowitz and Sorkin 1987) or *adaptive automation* (Rouse 1988). Although the rationale of this approach is plausible, it is not an easy task to determine which functions are to be allocated flexibly and which ones are to be allocated invariantly. Even more difficult is to decide who should have the ultimate decision power if there is a dispute between the system and the human. Applications of this are found in military aviation (e.g. Scerbo 1994) where the autopilot takes over if there are indications that the pilot has lost control (e.g. because of extreme centrifugal forces). In contrast to most analogues, the allocation of function in space travel may be considered a three-dimensional problem. This means that, in addition to the question of whether to allocate it to the human or the machine, one needs to decide whether the cabin crew or ground control would be better equipped to accomplish the function.

2.3.7 Operator support systems

Increasing complexity of technological systems have raised the demand for more support provided to the human operator helping him/her to take appropriate decisions. If the complexity of the system is high, it is difficult for the operator to find the 'right' data at the 'right' time (Woods 1986). This has been called the 'significance of data' problem. This is not caused by a lack of feedback, or informativeness, from the system (Bailey 1989), but rather the opposite when information overload makes it difficult to find the critical information. Being a common problem in many work environments, this has also been of concern during space missions where large amounts of data are to be processed (Connors et al 1985). Among the most successful measures to reduce the amount of information was the graphical presentation of data in the form of integrated spoke displays (Woods, Wise and Hanes 1981). The integrated graphical presentation of the information allows the monitoring of critical system parameters and deviations from target states are detectable at a glance.

During abnormal system states, two kinds of decision aids have been found to be very useful: *history displays* and *predictive displays*. In an emergency situation, when a large number of alarms are indicating a departure from a safe system state, the operator needs to know what were the preceding events that have led to the current situation. This historical information provided by an operator support system can be invaluable in determining the cause of an abnormal state. Similarly, with a large number of closely-coupled system variables to be considered, it is often very difficult for the operator (especially for the less experienced one who relies more strongly on closed-loop control) to understand fully all the interactions in the system so that the final result of the intervention can be accurately predicted. A predictive display, as part of the operator support system, may relieve him/her of some of the cognitive load by presenting him/her with the probable outcomes of the interventions considered.

Decision aids can be provided at several levels with regard to information presentation and operator control. They may be classified into three categories (Yoon and Hammer 1988): First, *information aids* that provide the operator with relevant information but do not give advice. Second, *expert systems* that give advice which is based on rules and procedures derived from a knowledge base, though the ultimate decision is taken by the operator. Third, *automatic diagnostic systems* in which the aiding system takes over most (or all) of the work by diagnosing and repairing the fault (possibly informing the operator about it) with all decision power lying in the hand of the machine. A debate has also evolved around the issue of whether an operator support system should be intelligent or not, because of its important implications (Duncan 1986, Reason 1990). Despite being very powerful, intelligent operator support systems also have a number of problems associated with them. First, they have cost implications because information aids tend to cost only a fraction of an automatic diagnostic systems. Second, operator involvement may be affected since automatic diagnostic systems tend to reduce it. Third, some decision support systems may threaten the expert status of the operator, hence not receiving sufficient operator acceptance. Finally, they were sometimes too difficult to use because of their high complexity. For example, Raaijmakers and Verduyn (1996) showed that even simple information aids (listing possible causes for system failures) can improve diagnostic performance substantially. Despite the importance of operator support systems in the context of automation, research has so far failed to carry out evaluations of the effectiveness of these systems in an applied context (Timmermans and Vlek 1992).

2.4 Human error in automation

A great deal of research into errors has been carried out because errors are a very common occurrence in many areas and may have costly consequences, in particular, in highly automated systems. While earlier work on errors in complex systems aimed at minimising errors, the emphasis has now shifted to minimise their negative effects (Morris and Rouse 1985). This has been achieved by increasing the detection probability of errors (improvement of feedback features of the system) and by making systems more error-tolerant or forgiving (incorrect actions can be reversed).

Although human error is considered to be a very frequent cause of accidents and of system breakdowns (compared to incidents attributed to machine failures), errors are often the result of poor system design or organisational deficiencies (Norman 1988). Accidents are usually caused by a multitude of factors which all contribute to the final outcome and none of the factors alone would have been able to cause the disaster. Some of these factors are active failures or *local triggers* while others are latent failures or *resident pathogens* (Reason 1990). The local trigger is the final event in a chain of events (e.g. poor system design, management failures) which may sometimes go back years before the actual accident. The resident pathogens, which represent previously committed errors without immediate consequence, remain dormant until activated by the local trigger - or until identified and removed before the occurrence of an incident. Many issues associated with automation (opaqueness, complexity, etc.) contribute to this (Reason 1990). In many ways human error is an unfortunate term since it has the connotation of the operator being entirely responsible for it. This often appears to be unjustified since the operator is only at the end of a long line of errors. In retrospect, with all the information available, it is easy to say: How could such an error have been made? However, in the actual situation, which is characterised by time pressure and the availability of only incomplete information, one would be unlikely to follow the same line of thinking and decision-making. Research evidence suggests that many people involved in accidents would not have been able to predict the final event from the pattern of previous events because the joint acting of several factors appeared to be unconnected at the time of the accident (Wagenaar and Groeneweg 1987). The authors called this the *impossible accident.*

A major problem in the context of automation is that, although it eliminates certain types of error, at the same time it introduces a different kind of error. Increased automation leads to a more frequent occurrence of higher-level operator errors (Wickens 1992; Reason 1990). In addition to the above mentioned local triggers and latent failures, there may be set-up errors, where the operator did not programme the automated device correctly before its use (Wickens 1992). A further error associated with automation is the mode error (Sarter and Woods 1995). They may occur in situations where the operator can choose among several modes of (automatic) control but selected the wrong mode for the sequence of action carried out.

An interesting approach for operator training would be error management (Frese 1996, Frese, Brodbeck, Heinbokel and Moser 1991). This approach has been developed in the context of office software and it has aimed to increase the positive effects of errors (e.g. refinement of mental model, more active system exploration) while reducing their negative impact (e.g. negative self-evaluation, anxiety). Although developed in a different operational environment, this approach may also be applied to the training of process control operators. While operators are encouraged to explore the system, the errors committed in the process are systematically discussed to improve the understanding of the system and operations are practised to recover the error. The results of the research showed that the suggested approach was advantageous, in particular, when it came to dealing with novel faults. However, the implementation of this approach in the process control industry, or even in spaceflight, may face considerable resistance because it may raise fears that too much of a positive attitude towards errors will cause operators to adopt too risky behavioural strategies.

2.5 Mental workload

There is an interesting relationship between mental workload and automation. The introduction of automation often results in a reduction of workload levels, and in some instances, automation is even introduced with the goal of reducing workload (Wickens 1992). This is primarily achieved during normal situations (i.e. when all operations run smoothly) but during abnormal situations workload levels can become excessive. Abnormal situations are often characterised by the unavailability of the automatic devices, that is, when automation is needed most, it is not available (Bainbridge 1987). Bainbridge described this as one of the *ironies of automation*. Furthermore, it was often found that when workload levels were reduced as a result of automation, they were subsequently increased again by giving new tasks to the operator (i.e. tasks that were previously completed by another person) to achieve productivity gains. As a result, there has been a continual reduction in staffing levels needed to run systems in various applications, such as in aviation (Wiener 1988) and in the maritime industry (National Research Council 1990).

Before implementing a new or redesigned system, one should assess the implications for workload levels. In practice, this is quite a difficult undertaking because of the complexity of the relationship between performance and mental workload. Mental workload assessment can be carried out with a number of techniques. Since they are used to assess the effects of sub-optimal working conditions and operator states, they represent a central part of the methodology of this thesis. They are discussed in detail in the following chapter, but are covered briefly here. Workload assessment techniques may be summarised under 4 headings: (1) Primary task techniques, (2) Secondary task techniques, (3) Physiological measures and (4) Subjective measures. Since the measurement of primary task performance has often failed to demonstrate effects of workload, some authors suggested the examination of primary task strategies (i.e. how is the task done rather than how well it is done), which appear to be much more sensitive to variations in mental workload (Spérandio 1995).

According to O'Donnell and Eggemeier (1986), a good workload assessment technique should satisfy five criteria. First, it should be sensitive to changes in task demands. Second, it should be diagnostic in that it indicates which capacity or resource is required by a given task. Third, it should be selective, that is, it should not be affected by factors unrelated to task demands (e.g. emotional stress). Fourth, it should be unobtrusive and not interfere with task performance. Fifth, it should be reliable over time. However, in practice most workload assessment techniques fall short of one or several of these criteria. Furthermore, when mental workload is measured with more than one technique, an inconsistent picture is sometimes found (Wickens 1992). This is because the techniques measure different aspects of workload. For example, subjective measures appear to be affected by effort expenditure but changes in effort are not necessarily reflected in performance (Yeh and Wickens 1988). Sometimes even parameters measured with the same technique, such as heart rate and heart rate variability, showed some degree of dissociation (e.g. Tattersall and Foord 1996). One may conclude that, the more techniques are used and the more parameters are measured, the more complete the picture will be. However, in practice, this often is not feasible because of problems of task interference, in particular, in applied settings.

Despite the difficulties in using workload assessment technique and the inconsistencies between them, workload assessment is an indispensable element in the system design process. Indeed, one may expect workload assessment to gain in importance as a result of increased complexity and automation, which make a priori judgements (i.e. without empirical evaluation) about the quality of the system design even more difficult and riskier. The assessment of workload in space is particularly difficult because of the many distinct mission phases, which vary considerably in terms of crew activities and demand levels.

2.6 Conclusion

The review of issues related to automation was of benefit in that it improved our understanding of the challenges faced in the design of systems for spaceflight. Generally, one can make good use of the knowledge gained from analogue work environments. In spite of the considerable degree of overlap between the relevant issues in spaceflight and most analogue work environments, their requirements differ in some areas because of the particular features of the space environment.

The first question concerns the implications for system design resulting from the fact that the ability and skill levels of space crews tend to be higher than those of operators in most terrestrial analogue environments. One may assume that confidence levels might be higher among astronauts, which may result in a stronger propensity towards manual intervention and hence a higher need for control (see Lee and Moray 1994). These differences may also lead to the development of a different, possibly superior, mental model of the different systems (DeKleer and Brown 1983). The question of training and the development of mental models is treated in more detail in Chapter 6.

A further question concerns the issue of system monitoring and allocation of functions. Functions need to be assigned, variantly or invariantly, between the machine, the cabin crew and mission control. Owing to the limited size of the cabin crew, there is a natural tendency to reduce their workload and assign to them only the functions that the other parties (i.e. the machine and ground control) cannot sensibly fulfill. This however may place the crew 'out-of-the-loop'. Furthermore, a 24-hour monitoring of space systems cannot be carried out solely by the cabin crew (unless work schedule arrangements are fundamentally changed, which is unlikely to happen before the crew and the cabin size significantly increase). The provision of adequate historical information about the system state and the external environment is therefore of primary importance to allow the crew to understand the sequence of events that has led to the current state.

Particular attention also needs to be paid to the those errors that appear to be a by-product of automation. For example, in-flight training on modern, high-fidelity computer simulations is considered a promising measure to shorten pre-flight preparation and to combat loss of skill as a result of an extended lay-off period. However, increased similarity between the training system and the operational system increases the likelihood of a mode error (i.e. training mode is mixed up with operational mode).

It appears that changes in workload levels have become more extreme as a result of automation. In extreme cases this may result in underload under normal operational conditions and exceptional levels of workload in emergency situations. This may be amplified in spaceflight by the different mission phases, which are associated with different workload levels. Considering this and the fact that back-up personnel is limited in space travel, system designers need to accommodate for possible extreme variations in workload levels.

The issues identified in this chapter are also of importance in guiding the design of the experimental task. Since this thesis aims to investigate human performance in complex work environments, the experimental task needs to contain the essential features of the reference environment. Ideally, it should simulate an automatic process, which however requires operator intervention at certain times. The process needs a certain degree of opaqueness, extensive system monitoring is required and operator support systems in some form need to be available.

It is obvious that only a few of these issues can be empirically tested in the form of independent variables in the experimental work. This is because they represent only those performance shaping factors that are associated with the design of the interface (see Chapter 3 for a more detailed discussion). For this thesis, it is planned to include a wider range of performance shaping factors, which also includes operator state and factors of the external environment.

Chapter 3:

Methodology

3.0 Summary

The thesis adopts a methodological approach which is firmly embedded in the tradition of human factors research. The methodology not only examines human performance but also collects measures of subjective operator state and strategies employed to achieve task goals. A model of compensatory control is used to guide the research and interpret the results. The model describes various adaptive strategies used by the operator to cope with changes in work demands. Some of the strategies represent hidden forms of skill decrement and, consequently, are not easily detectable. The research is conducted within the multiple-task paradigm, which acknowledges the complexity of behaviour in real work environments and the existence of sometimes conflicting demands of different task goals. Six experiments (3 lab-based, 3 field-based) were carried out, based on simulations of sub-optimal working conditions and sub-optimal operator states typically found in extended spaceflight. The independent variables were selected from three domains of human-machine interaction (operator task, human-machine interface, environment and operator state) to cover the key aspects of the space work environment. The analysis of the data is based on a normative approach, assessing the effects of different combinations of independent variables at three levels. First, performance on primary and secondary tasks is examined. Second, an analysis is carried out to investigate the strategies used to monitor and control the system. Third, the subjective response to these experimental conditions is measured by capturing levels of fatigue, anxiety and mental effort.

3.1 Theoretical framework

3.1.1 The human factors approach

As pointed out in Chapter 1, the thesis aims to improve our knowledge of human skill maintenance in complex and highly-automated work environments, with a particular focus on extended spaceflight. In order to achieve that goal, the methodological framework is based on the human factors approach, which aims to study human performance in work-related tasks within human-machine systems (Meister 1989). This does not imply an exclusive focus on performance but also examines the way task goals are accomplished. The human factors approach also makes use of physiological indices (cardiovascular, respiratory, nervous and sensory activities, and blood chemistry) to measure strain during task completion as well as subjective responses (perceived workload, comfort etc.) of the participants (Sanders and McCormick 1987). The tasks used for performance assessment need to be sufficiently sensitive so that increased effort, superior knowledge, more practice etc. result in better performance on the task. In Norman and Bobrow's (1975) terms, the task should be data limited during task execution, allowing us to distinguish between different levels of human-machine effectiveness.

Two different approaches to research in human factors may be differentiated. The first is concerned with generic research into the information processing capacity of humans, in order to develop guidelines to be used in different contexts. The second is concerned with the evaluation of a particular system in order to increase its efficiency and effectiveness. The research conducted for this thesis is part of the first strand in that it uses a generic simulation of a complex task to develop an understanding of human behaviour under different working conditions.

3.1.2 Stress and performance

Few people would dispute that space is a stressful working (and living) environment, given its many inherent stressors (e.g. noise, confinement, micro-gravity, danger). Further sources of stress may be found in a number of factors such as work tasks, organisational factors but also in individual attributes (e.g. tension with colleagues). To obtain a better understanding of the effects of these stressors, we need to look at the relationship between stress and performance. In many definitions of the concept of *stress*, the emphasis is on a mismatch of demands and available resources (e.g., Smith 1987, Greif 1991). However, stress is not only the result of environmental demands and characteristics of the individual, it also reflects a complex and dynamic interaction between the individual and the environment. This dynamic process makes it very difficult to predict the effects of given environmental demands on an individual. Despite interindividual differences in response patterns, a number of stressors (such as those mentioned at the beginning of the section) are generally considered to have a substantial impact on people's work behaviour.

The effects of work stressors on cognitive activities tend to show a highly complex pattern. First, the effect of different stressors on the same cognitive task can be very diverse. Sleep deprivation, noise, heat and physical confinement are all commonly regarded as stressors but their effects on the same cognitive activity may vary enormously (Hockey and Hamilton 1983). Second, the same stressor may have differential effects on different cognitive tasks. For example, while exposure to noise is likely to have a negative impact on a complex verbal reasoning task, it may improve performance on a simple checking task. Third, the joint presence of more than one stressor may have a qualitatively different effect than the effect of each stressor separately. For example, sleep deprivation and noise as individual stressors tend to result in performance decrements on certain tasks. However, when jointly present, sleep deprivation tend to reduce the negative effects of noise (Wilkinson 1963). Fourth, different aspects of performance (such as speed and accuracy) may be affected selectively (e.g. noise exposure tends to result in more errors while speed is often unimpaired). Fifth, the underlying mechanism of stress is often of high complexity. For example, increased stress levels have been found to lead to increased attentional selectivity, which may then result in performance decrements (Hockey and Hamilton 1983). This is because peripheral information sources or low priority tasks are increasingly neglected. The phenomenon of attentional narrowing may become more pronounced with an increasing number of tasks to attend to, which makes it interesting in the context of the multiple task environment used in the present case.

The impact of stress may be observed at several levels such as changes in physiological parameters, subjective state and performance. Physiological measures used for this purpose were, among others, heart rate (or heart rate variability), blood pressure and skin conductance, but also breathing patterns (Wientjes 1992) and hormones (Frankenhaeuser 1989) such as catecholamines and corticosteroids. Subjective state measures aimed to capture the individual's affective response (e.g. anxiety, anger, frustration) to stress-inducing situations. Performance may be affected in various ways as a result of stress, given that response patterns are qualitatively very different (see above). The effects of stress may manifest themselves very selectively, often not readily observable in the context of work performance. Many studies in stress research have adopted a too narrow focus by examining stress only at one level (e.g. physiological states or performance deterioration). This failed to address the complex pattern of the stress response. A more holistic approach is therefore needed, which embodies the different patterns of manifestation. This approach is pursued in this thesis.

3.1.3 The model of cognitive-energetical control mechanisms

The methodological framework in this thesis is based on a model of compensatory control to regulate performance under stress and high workload (e.g., Hockey 1993). Describing adaptive strategies in a framework of cognitive-energetical control mechanisms, it distinguishes between different patterns of degradation under stress and high work demands. (1) A decrement in *primary task performance* may be observed. However, this is rather rare in work contexts since the operator usually attempts to maintain primary performance levels because of the high costs associated with a severe performance breakdown. (2) Instead, one may observe various forms of *strategic adjustment.* Work goals may be lowered to avoid a catastrophic breakdown of performance or a less effortful strategy may be employed to conserve cognitive-energetical resources. This may manifest itself in decrements on secondary task performance while performance on high priority tasks generally remains unimpaired. (3) Rather than accepting a decrement in non-priority tasks, the human operator may attempt to prevent any impairment of task performance by an increase in effort expenditure. This response pattern is called *compensatory costs* since it implies psycho-physiological costs for the human organism in the form of a neuro-physiological response (e.g. release of catecholamines). This is likely to have an adverse effect on health and well-being if it is sustained. (4) A fourth pattern of decrement takes the form of *fatigue after-effects*, which may result from prolonged periods of sustained work. They may be detected only on sensitive probe tasks given at the end of a work session (e.g. in the form of low-effort strategies being employed).

Hockey's model does not attempt to make predictions about the detailed pattern of degradation under given circumstances. This is because the extent and focus of compensatory control is determined largely by factors under the control of the individual. It nevertheless makes general predictions of which tasks or task features are more likely to suffer under stress. The model is considered a useful framework for guiding the present set of studies. It is hoped that the experiments carried out in this thesis may help refine the model to improve its predictive properties. Although it is acknowledged that physiological parameters play an important part in the understanding of human reactions to stress, various constraints (lack of laboratory facilities and appropriate equipment, obtrusiveness of some of the measurement procedures) did not allow us to include any measures of this kind in the experiments.

The methodological framework adopted also stresses the role of the operator as an active agent rather than a passive respondent to environmental stimuli. According to action theory (e.g., Frese and Zapf 1994), any action completed by the operator is to be seen as goal-oriented behaviour, which is influenced by motivational factors. This is important for an appropriate evaluation of work performance because, only if the individual accepts the (usually) externally imposed work goals, and has sufficient motivation to pursue them, will we be able to evaluate any changes in work behaviour in the context described above.

3.1.4 Resource models

In the context of the cognitive-energetical model, the concept of *resources* plays a key role. It may be considered as the substance that fuels the cognitive process. A number of models have been proposed to describe the relationship between resources and performance in terms of performance-resource functions (see Norman and Bobrow's 1975). All models assume that resources are a scarce commodity and they can be allocated between tasks according to operational requirements. A distinction may be made between single-resource models and multiple-resource models. Single-resource models (e.g. Kahneman 1973) propose a single resource pool. Multiple-resource models (Wickens 1992; Navon and Gopher 1979) assume several resource pools, which allows time-sharing of certain tasks without a cost of concurrence. The evidence in the research literature appears to be in favour of the multiple-resource models, though the findings are equivocal. The weakness of these models is that they ignore the existence of an energetical component, which is under the independent control of the individual and can be activated in the event of extreme need (in Kahneman's model only physiological arousal is considered to increase the resource supply). This energetical component may take the form of an adaptive biological response, which causes the individual to release extra resources to deal with emergencies of various kinds. Motivational aspects of behaviour in resource allocation may also come into play (see action theory), resulting in an activation of the energetical resource pool.

Although originally developed in the context of human error, Rasmussen's (1986) model may also be applied in the context of resource expenditure. The level of cognitive processing (skill-based, rule-based or knowledge-based activities) largely predicts the consumption of resources. This has important implications. For example, it implies that effects of training are twofold. First, it improves performance directly through increased practice and improved knowledge. Second, it leads to a reduction in resource consumption since lower processing levels (e.g. skill-based) are used for task execution, which require fewer resources. Rasmussen's model represents an important element in the rationale of the experiments conducted.

3.1.5 Multiple-task paradigm

The multiple-task paradigm has been adopted in this thesis to obtain a better understanding of the management of multiple demands in a real work environment. This was because a dominant feature of most work settings is the simultaneous pursuit of several tasks. A considerable body of research has used the multiple-task paradigm (see Proctor and Dutta 1995). In most cases it is more accurately described as the dual-task paradigm since few research studies have employed more than two tasks.

A modification from the traditional secondary task technique has been used in the form of a *loading task* (Ogden, Levine and Eisner 1979). In this technique the participant is asked to expend all resources on the secondary task (= loading task) while performance on the primary task is measured as an indicator of task interference. This approach appears to be of little use in the present study since it would contradict intrinsic task priorities.

Most studies employed artificial secondary tasks with little ecological validity. Typical secondary tasks employed were rhythmic tapping (Michon 1966), reacting to a probe stimulus (Posner and Boies 1971), and time estimation (Hart 1975). Although these kinds of tasks may be useful in a lab-based setting to measure cognitive load, they are likely to face user acceptance problems in real work contexts due to their low ecological validity. The tasks used in this thesis were designed to be more meaningful in the operational context of spaceflight.

Besides higher ecological validity, a further argument can be put forward in favour of a multiple-task environment containing more than one secondary task. The same secondary task may vary in terms of its sensitivity to high work demands as a function of the degree to which it competes for the same resources like the primary task. Kahneman (1973) therefore proposed to use a battery of secondary task measures as the ideal secondary task technique.

3.2 Method

Many methodological approaches have been employed to investigate human behaviour in a process control environment. Studying the issue in a real work environment has often faced the problem of not having sufficient control over intervening variables, which are an inherent part of any natural setting. On the other hand, laboratory studies often lacked the complexity of the real work environment by focusing on isolated aspects of work (e.g. STM, monitoring). Even if dual tasks were used they often comprised two artificially combined tasks (e.g. tracking and counting backwards) with little ecological validity. The rationale behind these methodological approaches has been the exploration of residual mental workload capacities by using different cognitive tasks (see Wickens 1992). Results obtained with these methods are valuable but that approach fails to address pertinent features of multiple task management in real work environments (such as complexity, and conflicting and changing task goals).

One of the problems has been the difficulty of developing complex simulations of real-world processes for use in laboratory experiments. Recently, with advances in computer technology, it has been possible to reduce the gap between lab-based experiments and field studies. Being able to simulate complex processes on a personal computer has enabled researchers to bring into the laboratory more realistic task scenarios, based on complex work environments with dynamic features (Brehmer, Leplat and Rasmussen 1991). A number of computer-based simulations (e.g., DESSY, Lohhausen, Pasteuriser; see Chapter 4) have been used in research to improve our understanding of human cognition in dynamic task environments. This approach has also been adopted in this thesis. AMT (air management task) has been developed to have a simulation task available, which is tailored to the needs of the research programme (a detailed description of AMT is given in Chapter 4). Using a realistic scenario in a generic simulation task provides a large amount of data about many aspects of process control behaviour. AMT is suitable for lab-based research as well as in a work-based context, given the high level of face validity of the simulation task.

3.2.1 The experiments

To address the issues outlined above, a series of six experiments has been carried out. Three of these were conducted in a laboratory setting and three in a field setting. The issues addressed in the lab-based studies differed slightly from those in the field studies. In the lab studies a number of independent variables were used to simulate different configurations of difficult working conditions, with the goal of assessing their effect on a range of cognitive activities. Between 16 and 25 participants were selected for each experiment. Participants were given extensive training on the task so that they would reach asymptotic levels of skill and performance. All participants had a degree (or were close to obtaining one) in engineering or a science. This was to ensure that the profile of the participants had some degree of similarity with the participants used in the field experiments.

The field studies provided the possibility to observe changes in cognitive task performance in a more realistic work environment while using highly-qualified personnel. The dependent measures were primarily analysed as a function of time spent in an isolation and confinement environment but also included other study variables (e.g. control load, fault type). It was planned to achieve this by taking measures in the pre- and post-isolation period, with the beginning and the end of the isolation period marking natural points of interventions, which may affect the outcome variables. The number of participants was smaller than for the lab-based experiments (between 3 and 16 in each experiment).

Experiment 1

Since the lab-based experiments did not allow us to simulate isolation and confinement, the emphasis was on a generic simulation of some characteristics of the space environment, which are expected or known to impinge on crew performance. Partial (occasionally also total) loss of sleep is a common occurrence in spaceflight. Such problems can sometimes accumulate over longer periods of time. The effects of sleep deprivation were therefore addressed in the first lab-based experiment in combination with task difficulty and dialogue control (as a possible moderator variable) to ascertain which cognitive activities suffer most under these sub-optimal working conditions.

Experiment 2

This experiment examined the effectiveness of certain training regimes under different operational conditions. The participants were presented with different types of system failures (e.g. routine, novel and unrepairable). In addition, noise was included as an ambient feature of space travel (continuous noise is caused by the on-board life support systems) to assess its impact on different aspects of performance. This experiment required a substantial modification of the simulation software, which placed stronger emphasis on diagnostic behaviour by increasing the number and complexity of simulated system failures. It also provided a history display as a decision aid to support operator decision-making (see Chapter 4).

Experiment 3

This experiment investigated the question of retention of skill over an extended interval. This is an important issue in extended spaceflight for low-frequency skills (see Chapter 1). Participants taking part in Experiment 2 were invited back to the laboratory for re-testing 8 months after their first testing session. This allowed us to test the stability of these complex skills during an extended lay-off period. Retention was studied as a function of the different training approaches adopted in Experiment 2.

Experiment 4

This field study looked at the effects of short-term isolation and confinement, involving four Canadian astronauts, who were isolated and confined in the mock-up of a space capsule for a week. Performance was assessed daily by looking at diagnostic skills and manual control skills required for controlling a technical system of this kind. Comparisons were made between different mission phases (pre-, in- and post-mission phase) and levels of control load.

Experiment 5

The second field study is closely related to Experiment 4. Using the same simulation tool, it addressed the same research questions but looked at them in the context of long-term isolation and confinement. The experiment took place in a French over-wintering camp in the Antarctic and lasted for 8 months. Testing intervals were longer than in experiment 4 because of constraints of the study.

Experiment 6

In the final study, the emphasis was on the effects of medium-term isolation and confinement. The study took place during a 135-day simulation of an orbital stay on the Russian MIR-station, involving frequent testing of the three cosmonauts with the AMT research tool. The ecological validity of this experiment was higher than in the two others since the technical as well as the ambient environment (prolonged isolation) was very similar to real extended spaceflight. The independent variables looked at were mission phase, isolation and confinement and fault type.

3.2.2 Selection of experimental variables

The selection of experimental variables was largely driven by the relevance of research questions, which needed to be addressed in the context of extended spaceflight. Furthermore, two experiments (No 1 and 2) were based on research grants sponsored by the European Space Agency, which meant that the design of the experiments had to take into account the interests of the contract sponsor.

The experimental variables were classified into three domains: (1) Operator task, (2) Human-machine interface and (3) Environment and operator state. This distinction was based on a classification system proposed by Roth and Woods (1988), which they called the *cognitive triad.* The cognitive triad proved to be a useful classification to guide research into human-machine interaction. The first domain *operator task* refers to the cognitive demands imposed by the world onto the operator. The second domain *human-machine interface* is concerned with the operator's external representations of the world, which are significantly influenced by the design of the interface with the machine. The third domain is called *environment and operator state* to indicate the capacities of the operator as a function of external environmental agents (e.g. noise, heat) and knowledge of different kind (e.g. special training). All three domains contribute to the overall effectiveness of human-machine systems, with mismatches in any of them leading to non-optimal performance.

Table 3.1 summarises all the independent variables examined in this thesis, indicating how they fit into the cognitive triad. The selection of variables indicates the emphasis on *operator task* and on *environmental and operational state* factors as the main thrust of the research. The domain *human-machine interface* is more concerned with issues at a micro-ergonomic level such as information display and controls.

Operator task:	• Workload/task difficulty/fault type (1-6)
Human-machine interface:	• Dialogue control (1)
Environment and operator state:	• Sleep deprivation (1) • Isolation and confinement (4,5,6) • Noise (2) • Training (2,3) • Skill retention (3)

Table 3. 1: Three domains of cognitive triad (based on Roth and Woods 1988) with all independent variables that were employed in this thesis (The experiments researching the independent variable are given in parenthesis)

3.2.3 Levels of analysis

Corresponding to the conceptual framework outlined earlier, the main methodological approach was based on a normative analysis. It aimed to assess the effects of different combinations of the experimental variables at three levels: (1) Performance on primary and secondary tasks, (2) Control and information sampling behaviour, and (3) Subjective state.

Performance on primary and secondary tasks

The complexity of the task environment (with partly conflicting goals) has the advantage of being able to observe possible differential effects on the different tasks as a result of their differing cognitive demands and different levels of priority attached. As already pointed out above, the secondary task technique has proved to be a very useful method to assess primary task demands.

Control and information sampling behaviour

Changes in operator control behaviour are not necessarily reflected in performance measures (e.g. Dörner and Pfeifer 1993). For example, two completely different diagnostic strategies (e.g. 'trial-and-error' method vs 'reflection') may be alike regarding certain performance criteria (e.g. fault rectification time) even though they represent fundamentally different diagnostic approaches. It would be an important observation if an operator switched from a systematic approach of fault diagnosis (high resource demands) to a 'trial-and-error' approach (low resource demands) under increasing task demands.

Subjective state

The third level of analysis measures operator perception of the task in terms of demand and psychological comfort. They are not always found to be associated with performance (Yeh and Wickens 1988). For example, when performance maintenance was obtained only at the cost of increasing effort expenditure to the upper limit, it has led to the depletion of internal resources. As a result, effort ratings were increased while no effect on performance was found. This makes subjective operator state an interesting complement to the other measures.

3.2.4 Performance variables: primary and secondary tasks

In accordance with the multiple-task paradigm, a number of tasks were selected and incorporated into the overall task environment. They differ regarding their sensitivity to stressors of different kind and are therefore useful indicators of the impact of different configurations of sub-optimal working conditions. This section provides a brief summary of the main performance measures used and how they were scored.

System control performance. This is one of the primary tasks of a process control operator and its successful completion is based on a number of different factors (e.g. maintaining vigilance, system knowledge, information sampling behaviour). The factors determining system control performance are discussed in relevant sections of the thesis (see Chapters 2 and 6). System control performance was measured by recording the time during which any system parameter was not in its target state, resulting in an error score (parameter control failures). This measure tended to be positively skewed, since the more highly skilled participants showed scores close to zero whereas the poorer operators' scores were spread over a larger range at the other end of the distribution.

Fault diagnosis. This is also a primary task for a process control operator since after successful system stabilisation, efforts are required to identify the cause of the system disturbance. Again, a considerable body of research has focussed on various aspects of diagnostic performance. In three experiments a distinction was made between diagnostic speed and diagnostic accuracy to be able to identify any possible speed-accuracy trade-offs. In the other three experiments only diagnostic accuracy was measured. Diagnostic speed was scored by recording the time elapsed between the beginning of a system fault and its reporting. The data tended to show only a slight positive skewness. Diagnostic accuracy was scored by counting the number of incorrect fault diagnoses. Skewness tended to be quite strong.

Reaction time. This secondary task provides an indicator of cognitive load by measuring the time needed to respond to a warning signal (e.g. Posner 1986). RT is valuable as an indicator of primary task load in the context of secondary task techniques (see Wickens 1992 for an overview). Each system parameter was provided with a specific warning signal to indicate that the parameter was out of range. False alarms were designed to be relatively frequent (≈ 1/min) to provide a reliable index of RT for each session. The data distributions were close to normality and did not require any transformations.

Prospective memory. This refers to the ability to remember to carry out an action in the future. The research literature in this area is comparatively small, though interest is increasing. Prospective memory has been studied in contexts such as ageing (Cockburn and Smith 1994) and in clinical settings with patients suffering from severe impairment of prospective memory as a result of head injury (Mateer et al. 1987). Other strands of research were concerned with the strategies that may improve prospective memory (Morris 1992, Cohen 1989). There seems to be evidence that demands for this task undergo phasic changes with time monitoring activities being highest at the end of the time period (i.e. when some action is required), while considerable fewer checks of the clock occur in the middle of the period (Harris and Wilkins 1982). Ellis (1988) made an interesting distinction between different kinds of prospective memory: *Pulses* and *Steps*. Pulses are activities that need to be done at a fixed time (e.g. being at the doctor's at 9.45) whereas steps can be completed during a time period (e.g. visiting a friend by Friday). Prospective memory appears not to have been studied in work settings. However, there is some evidence that errors in process control applications sometimes occur because of operators not remembering to carry out a certain action (e.g. the closure of a valve) at a fixed time (Reason 1990). Requirements in process control would include both, pulses and steps. The following scoring system was used for the prospective memory task. Recordings were scored for accuracy of timing and assigned to four categories: Early (-5 sec or more), correct (+/- 5 sec), late (+5 sec to +20 sec) or omissions (> +21 sec or missed). The number of recordings in each category was transformed into a percentage of total recordings. The proportion of omitted recordings was then analysed by analysis of variance. This category represented the strongest indicator of prospective memory failures.

Resource consumption (control efficiency). This measure is closely related to system control performance, representing essentially a second order goal during control activities. It puts certain constraints onto the operator since selected control strategies (providing there is a choice) also need to be checked for compliance with efficiency criteria. Due to these constraints it is regarded as a difficult task since the need for immediate action in the primary task may not allow for careful considerations of the impact of every possible action on the resource reserves. The control efficiency measure is the equivalent of a *cost factor* that is considered as an important element in real-life decision-making (Brehmer 1992). Performance was measured by calculating the percentage of resources (O_2, N_2 and energy) consumed during a testing session.

3.2.5 Control and information sampling behaviour: operator strategies

This level of analysis considers the strategies used to manage the system under different working conditions. The measures provide an indicator of performance efficiency, as opposed to performance effectiveness (see Hockey 1996). Although the relationship between the frequency of an action and its effectiveness is by no means easy to establish, a frequency measure may be considered as an indicator of changes in activity levels or biases towards certain control modes (e.g. from feedback to feedforward control).

Control action. Effective control actions are the basis for skilled system management. They may take the form of manual control actions or adjustments of automatic control parameters (auto trimmings). The latter require less system knowledge and are easier to accomplish. Measures were scored separately for different subsystems and types of control actions by counting the number of interventions per unit of time.

Monitoring actions. Information sampling behaviour is treated in close tandem with control actions, since together they provide the basis for an effective intervention strategy. Again, the frequency of information sampling is not necessarily an indicator of the quality of performance but it may indicate adaptive changes in sampling strategies. Monitoring actions were scored by counting the number of times operators sampled the display during the session.

3.2.6 Subjective state measures

In order to assess the subjective state of the operator, a modified version of SWAT (subjective workload assessment technique) was used (Reid and Nygren 1988). This self-assessment technique has been regarded as sufficiently sensitive to changes in task demands. Furthermore, it is unobtrusive (unlike concurrent workload evaluation techniques, see Tattersall and Foord 1996) and easy to administer. Rather than using a three-point scale as in SWAT, a 100 mm visual analogue scale was used to capture the subjective operator state.

Whereas SWAT measures three dimensions of workload (time load, mental effort load and stress load), this thesis collected measures of *mental effort*, *anxiety* and *fatigue*. While the subjective workload items were collapsed into a single index (mental effort), it was considered important to include an indicator of perceived threat of environmental demands (anxiety) and an indicator of the consequences of continuous task involvement (fatigue). The limitation of the subjective state indicators to three items reflected our concern to minimise disruptions associated with the data collection process, in particular, as these were measured repeatedly during task completion.

3.2.7 Data analysis

Analysis of variance was used as the main statistical method for data analysis. Although analysis of variance is very robust against violations of the assumptions of normality and homogeneity of variance (Howell 1992), in some cases the data needed to be modified by a square-root or logarithmic transformation. This became necessary because many of the dependent variables had a positively skewed distribution. This is very common with performance measures, where zero values represent optimal performance and deviations from zero indicate non-optimal performance (errors). Therefore, the scores of the poorly performing participants are distributed over a wide range whereas the scores of the good participants are concentrated close to zero. However, for the sake of clarity all tables and graphs found in the thesis always contain the untransformed data.

The design constraints in the field experiments (Chapter 7) caused some methodological problems with the use of analysis of variance. They used repeated measurement designs with only small numbers in each group (except for Experiment 5). Nevertheless, it was still possible to use analysis of variance to identify some overall group effects, but the power of the test was small. As a first step, visual inspection of the data is required to detect any obvious trends. The suggestion has been made by some authors that visual inspection as the sole analysis may be sufficient in some circumstances (usually clinical settings) to draw conclusion about the effect of the interventions (e.g. Wilson 1995). However, it was not judged to be suitable in the present case to base the analysis entirely on this approach, although it is acknowledged that in certain circumstances this might be the only option.

3.2.8 Participant selection

In the lab-based studies, it was intended to use a participant pool that showed some similarity to the profile of real space crews. This was achieved through the use of some basic selection criteria, which avoid some of the generalisability problems commonly associated with the use of undergraduate psychology students. It was therefore required that the study participants had a degree in science or engineering and an adequate level of computer literacy. Students in the last year of their first degree were also regarded as acceptable for the studies. Participant selection for the field studies was not under the control of the experimenter. Selection for these studies was carried out directly by the

European Space Agency and consisted of a detailed and elaborate procedure (psychometric tests, interviews etc.).

Chapter 4:

Simulation Task

4.0 Summary

There is an immense potential for computer-based simulations of task environments, allowing research to be carried out into relevant issues in complex work environments without facing many of the problems associated with research in real operational environments. Having reviewed a number of existing task simulations, it became evident that they were not suitable for our purposes. This resulted in the development of our own task environment - named AMT (Air Management Task) - which had some distinct advantages over the existing simulation packages. First, it was based on a multiple-task methodology, which is an essential part of a high fidelity simulation of the real world. Second, it was a truly dynamic task unlike many of the simulation tasks reviewed. Third, it provided a higher level of face validity for the subjects in the field study, which is an important consideration where user acceptance and commitment is crucial. A detailed description of the different AMT versions is given.

4.1 Computer-based task simulations

4.1.1 Introduction

The rapid advances in computer technology, which resulted in the development of increasingly powerful machines, brought a great deal of benefit to ergonomists studying human behaviour in the context of complex human-machine systems. It enabled them to use PC-based simulations of complex systems in research laboratories, whereas previously it was necessary to use highly sophisticated mock-ups of real systems to address the same issues. The latest developments now offer the advantage of analysing performance in the complexity of real-life scenarios while conducting the research in a well-controlled experimental setting (Brehmer, Leplat and Rasmussen 1991; Brehmer 1992). This has the advantage of being able to confront subjects with the same task scenario, where all actions can be recorded (those from the operator as well as the system), allowing the study of structural relationships in greater detail. The second advantage is being able to study human performance in dynamic tasks, which is distinctly different from the traditional use of (somewhat unrealistic) static performance tests. This means that decisions have to be made in real time, where the simulation software does not stop and wait until the individual has taken a decision. A number of research groups have developed simulation software packages with the intention of creating a more realistic task environment while benefiting from the advantages outlined above. A brief review of the most widely used packages is given below.

4.1.2 Rationale for using computer simulations of complex processes

In this section the benefits of using realistic simulations of aspects of work environments shall be argued for, in order to better understand human behaviour in complex dynamic environments. For many decades, the single task method was predominant in experimental psychology research, where different components of cognition (e.g., vigilance, working memory, reaction time) were separately assessed under different experimental conditions. This basically represented a decoupling of cognitive components from each other although it has been widely accepted that the different components usually operate in interaction with each other (Eysenck 1984). Furthermore, it has been argued that the ecological validity of most tasks used in the classic experimental literature of cognition was very low since they hardly had any relevance to everyday life (Claxton 1980).

As a result of the shortcomings of this approach, a stronger tendency to employ multiple task methods was observed (Proctor and Dutta 1995). However, multiple task methods also largely employed static tasks, lacking some essential features of a dynamic task environment, which are essential characteristics of real-life scenarios. The following characteristics are fundamental components of a dynamic task environment (Brehmer 1992; Brehmer and Allard 1991): *Time*, *Complexity* and *Delays*.

Time. This is a crucial factor. Taking process control as an example for a dynamic environment, it requires continuous up-dating of the mental model of the current system state since the problem space is constantly changing even without operator intervention. This is because the system state is influenced by several agents such as the operator and the automatic controller. In addition, environmental factors (e.g., failures of system components, fall in outside temperature) have an impact on the system state.

Complexity. In a dynamic task environment, a large number of system interactions may cause frequent changes in the system state, of which the precise cause-effect relationships are sometimes, even for the expert operator, difficult to discern.

Delays. Any intervention carried out will normally show any effect in the system state with a time delay. It is widely acknowledged that the increasing time delays in a system have very negative effects on operator performance because of the difficulty of assessing the impact of one's own control actions onto the target state.

However, despite the new opportunities offered through computer simulations, there are also limitations to the simulation as an experimental tool (Brehmer et al 1991). First, as in any other experiment, the variables are selected by the experimenter, which influences the outcome of the study. The result may be that important effects are completely ignored because the corresponding variables were not included in the simulation software in the first place. This may have been due to the experimenter's belief (often influenced by the shared belief of the research community) that they are not of relevance in the research context. However, these are general methodological considerations, which do not solely apply to simulations but to any experimental activity. Second, the psychological research paradigm of a strict causation from stimuli to responses cannot be upheld any more since the stimuli are also produced by the operators themselves due to the interactive feature of the simulation (Brehmer 1992; Brehmer et al 1991). This means that during the experimental session the actual problem domain faced by the operator depends to a large extent on the preceding behaviour of the operator. Therefore, the independent variables should be regarded as system characteristics (e.g., delay of feedback, complexity, opaqueness).

In the context of political decision-making simulation programmes, computer-based simulations have been named *microworlds* (Brehmer 1992). This seems to be an appropriate metaphor for looking at the current and future potential of these tools. Although the use of a complex scenario raises some questions with regard to experimental control, its obvious benefits should be taken advantage of.

4.1.3 Review of existing simulation software

A number of computer-based simulations have already been developed to address the broad issue of human cognitive behaviour in complex and dynamic task environments. The most prominent simulation environments are now reviewed.

4.1.3.1 Process control simulation tasks

Waterbath. This is a process control task which was originally developed by Crossman and Cooke (1974) in a non-computerised form. Later it was changed into a computer-based simulation to carry out research in the microworld paradigm (e.g. Moray and Rotenberg 1989). It consists of 4 vessels filled with water which are connected to each other by a system of pipes controlled by a number of valves. The goal of the operator is to bring the system from a starting point to a given target state (with a given water temperature and fluid level) and maintain it there. The system operated with a considerable time lag between the heater being switched on and the attendant temperature changes. Furthermore, a number of system faults (e.g. valve blockages) can be set-up to simulate disturbed system states. The simulation task was truly dynamic (i.e. process proceeded even without operator intervention).

Pasteuriser. Originally developed by Muir and Huey as part of their PhDs, this is a simulation of an orange juice pasteurisation plant, which has already been used in a number of experiments (for an overview see Lee and Moray 1994). It comprises three subsystems (feedstock pump, heater and steam pump) which need to be controlled effectively to maximise the amount of orange juice produced. Each subsystem is controlled by an automatic controller but the operator is able to take over manual control at any time. Fault states can be set up that require the operator to adopt manual control. The simulation task has most characteristics of a process control system such as high complexity, time lag, interdependency between subsystems and most importantly it was truly dynamic. Despite its complexity, it was not a multiple task environment since all activities were concerned with parameter control, pursuing the overall goal of production maximisation (i.e. no conflicting goals). However, a second task (data logging) was sometimes used (but not measured) to increase overall task difficulty and encourage operators to make more use of the automatic control systems.

Plant. This is a computer-based process control task, where the operator has to supervise the flow of fluid through a network of tanks which are connected by valves (Morris, Rouse and Fath 1985). The goal of the operator is to maximise output within the limits of the production system. The system has its own dynamic interactions, which are independent from operator input but the task itself is operator-paced. This means that the system states are only updated after an operator intervention. The task requires the operator to diagnose, repair and compensate for system failures since the system does not operate with complete reliability. Four types of system failures may occur: Tank failure, pump failure, tank rupture and failure of safety system.

4.1.3.2 Non-process control simulation tasks

DESSY (Dynamic Environmental Simulation System). This task, which is sometimes also referred to as the *firefighting task*, was developed by Brehmer and his colleagues at Uppsala University in Sweden (see e.g. Brehmer and Allard 1991). It is essentially a managerial decision-making task, where the subject is in the position of the chief of a fire brigade with the task of coordinating a number of firefighting units (FFUs) to extinguish forest fires in the surrounding areas and to protect the base of the fire brigade. The fire chief is provided with information about the position and current activities (e.g. firefighting or inactive) of FFUs, the location of fires and weather conditions. All information is provided simultaneously on a computer screen. The simulation task allows the manipulation of a number of system characteristics: delay of feedback, complexity (varying effectiveness of FFUs) etc. The task is truly dynamic since the process will not come to a halt if the operator does not enter a command (the screen is up-dated in 20-sec steps).

Moro. This task was developed by Dörner (1990) and is essentially a managerial decision-making task, where the subject is in the position of an advisor to a tribe in a region in Africa called Moro. Decisions are required with the goal of improving living conditions for the indigenous population over a simulated period of 20-30 years. The task is complex, dynamic and opaque. The complexity exists because of the considerable number of variables (arable land, size of human population, size of grassland, tsetse fly population, cattle population, ground water, etc.), which are affected by various interventions (sinking wells, increasing arable land, etc.). It is truly dynamic (or a continuous simulation as Chubb, Laughery and Pritsker (1987) would call it) since the process is driven further even if no human intervention takes place. It is opaque in the sense that some variables (e.g. groundwater levels) are not directly observable. All three attributes are characteristics of real work environments.

Lohhausen. This task is in many aspects similar to *MORO* concerning the underlying dynamics of the task but it differs with regard to the operational environment (Dörner, Kreuzig, Reither and Stäudel 1994). With the participant being in the position of the mayor of a small town, it simulates many aspects of the decision-making requirements in the management of a community of this kind. Having nearly unrestricted power, the mayor has to take decisions in a large variety of areas, ranging from kindergarten and school policies to the management of a council-owned local factory. The system is truly dynamic since the process continues even in the absence of human intervention. The experimental task is usually completed over 8 sessions of 2 hours each, simulating a period of 10 years.

Joy. Having been especially developed for space research, this task has already been employed in a number of space-related research studies (e.g., Gushin, Efimov and Smirnova 1996). It contains nine tasks in the form of computer games (target shooting, finding a way out of a maze, etc.), covering a range of different cognitive activities such as eye-motor control, attention, working memory, mental arithmetic, spatial orientation, logical reasoning. The tasks are modular, allowing a flexible set-up for the purpose of the experiment. Performance measures of different kinds(accuracy and speed) are recorded at the end of each task (lasting between 1 and 3 min). Although being especially developed for space-related research, the software consists of single-task modules which do not have any greater relevance to space than to any other complex work environment. Furthermore, the task lacks any interactive or dynamic features, which are part of all complex real-world systems of this kind. Despite these shortcomings, it has been included in this review because it has already been extensively applied in space and space-related research.

4.1.3.3 Conclusion from review of task environments

The review demonstrated that there are a number of task environments available of which some (e.g. DESSY, Pasteuriser) have been extensively used in research. They cover a wide range of different activities with most of them focusing on managerial decision-making and process control applications. There is considerable variation concerning levels of complexity with some task environments scoring very low on that dimension (e.g. Joy) whereas others (e.g. Lohhausen, Moro) are characterised by a considerable degree of complexity. Although most authors claimed that their simulation tasks were dynamic, only some had truly dynamic features in that the system continues to change its state even in the absence of operator intervention. This is a very important point since in virtually all real-world systems (be it aviation, process control, ship navigation or air traffic control), the environment and the process continue to change. This means that the problem domain faced by the operator in the decision-making process is strongly time-dependent.

If a simulation task is described as complex in this context, it is by no means a reference to industrial standards. The perception of complexity is of a relative nature and is strongly influenced (among other factors) by operator training. Training in the context of a simulation exercise may entail perhaps up to 5 hours of practice and instructions, but it would last many months or possibly even years to manage a system in an industrial context. Therefore, due to the reduced training time in simulation studies, the task is likely to be sufficiently complex for the operator and may represent a serious challenge throughout the experiment.

Given the limitations of the simulation tasks available at the time, there were considerable benefits in developing a task that would incorporate the features previously identified as being essential characteristics of a process control environment. Furthermore, if the face validity of the task was high, this would increase user acceptance, particularly, with regard to real space crews. They would probably have been considerably less motivated if they had been asked to pasteurise orange juice throughout their mission, rather than using a relevant task which simulates (with considerable fidelity) aspects of real space work. This was also pointed out by Brehmer et al (1991) who asserted that the cover story in a computerised simulation task has a considerable impact on motivation and performance.

Concerning the fidelity of a simulation, AMT also meets three important criteria. These were identified by Wickens (1992) as features of real process control environments but are usually not found in simulations for lab-based vigilance studies: (1) Moderate operator activity levels between system failures result in the maintenance of at least modest arousal levels, (2) Alarms are sufficiently salient to attract operator attention and (3) An excessive number of false alarms occur during normal plant operation.

The main characteristics of AMT, which distinguish it from most other simulation tasks, are now summarised. (1) The task is truly dynamic, i.e. it is not operator-paced. (2) It has high face validity in the context of spaceflight. (3) The system is of considerable complexity. (4) The multiple-task environment makes it very realistic because of the different priorities attached to individual tasks.

4.2 Description of AMT

4.2.1 General features of AMT

4.2.1.1 Introduction

AMT is a PC-based simulation of a life support system in a space vessel. The life support system is highly automated and managed by a number of control devices, which keep the key parameters within their target state. The control devices, which are essentially subsystems of the life support system, are closely coupled, and therefore, have an effect on each other during system operation. While the system is running automatically, the task of the operator is to monitor the performance of the automatic controllers and to intervene (i.e., to take over manual control) if there seems to be a system fault that causes one or more of the primary parameters to leave their target state.

4.2.1.2 The key parameters

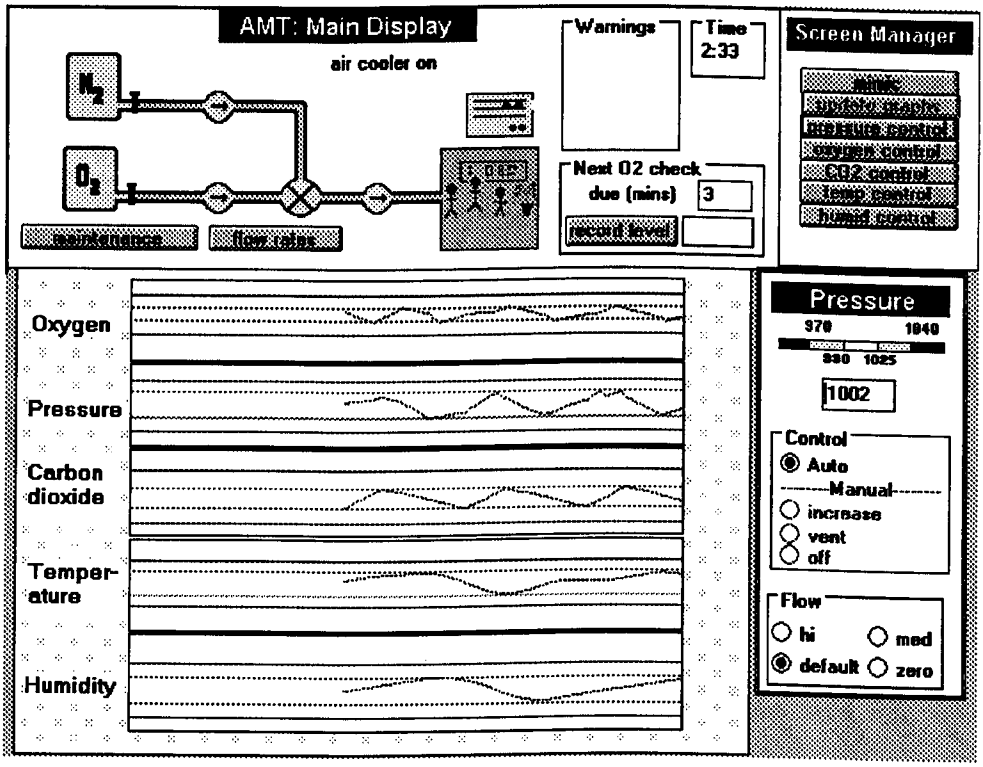
The main task of the operator is to monitor air quality in the space vessel by maintaining the following five parameters[1] within their target state.

Oxygen (19.0 - 20.5 %)
Pressure (970 - 1040 mbar)
Carbon dioxide (0.1 - 1.5 %)
Temperature (18.5 - 23.0° C)
Humidity (36 - 44 %)

In order to aid effective control of the parameters, three zones were defined, in which a parameter can move. The *white zone* defines the normal operating range, wherein the parameter should oscillate ideally. The automatic controllers have their set points at the boundaries of the white zone and therefore effectively maintain the parameter in that area. However, occasionally the parameter drifts outside that range into the *yellow zone*, before it is brought back into the white zone by the automatic system. If the deviation from the white zone is considerable, it can be an early warning sign of a failure of a system component. Then there would be an elevated risk of the parameter moving into the *red zone*, which is to be avoided. If a parameter moves into the red zone, it is counted as a parameter control failure. The ranges provided after the key parameters above refer to the boundaries of the red zone.

4.2.1.3 Interface of the operator

The main display provides a mimic of the topographical layout of the system with some additional features such as a warning system and a clock. The system consists of the following basic components, which are presented in figure 4.1: Nitrogen tank, Oxygen tank, Nitrogen flow meter, Oxygen flow meter, Mixer, Space cabin and Air conditioner.



Time 2:33

Next O2 check due (mins) 3

1002

Control: ● Auto

Flow: ● default

Figure 4. 1: Main display of AMT 3.0

For each parameter a control panel is available[2], enabling the operator to manipulate the levels by taking over manual control. At the top of the control panel, a coloured graphic indicates the different zones for the parameter in question. Below the zone indicator the current level of the parameter is displayed. Since the design of the remaining part of the control panel varies slightly between versions of the software, details are given in the corresponding sections.

The main display also provides some feedback about the operation of various system components. Flow meters indicate the flow of gases at several locations in the pipework. The icon of the mixer valve rotates when either gas is flowing. When the CO_2 scrubber or the air venting system is in operation, a small red arrow moves in and out. The operation of the cooler, the heater and the dehumidifier are indicated by the display of the appropriate message.

If any of the key parameters moves out of range, a red warning signal is displayed on the screen.

4.2.1.4 Basic mechanisms of the life support system

The cabin is supplied with a mixture of gases by the two tanks, a process which is controlled by two subsystems (pressure and oxygen) of which a more detailed description is given below. The air conditioner removes gases (O_2, N_2 and CO_2) from the cabin by the means of two subsystems (pressure and CO_2 scrubber). In order to maintain temperature at the prescribed levels, an evaporative cooler and a heater operate in tandem while being controlled by the temperature subsystem. Humidity levels are controlled by the dehumidifier, reducing the moisture content of the air, which steadily rises due to respiration and sweating.

Each subsystem operates by using set points, which trigger off some regulatory activity once a parameter has reached this prespecified level. Now each subsystem is looked at in detail.

Oxygen control: The oxygen control system maintains oxygen levels within a range of 19.5% to 20%. When the concentration of oxygen falls below this threshold, the O_2 supply is switched on and oxygen flows into the cabin. When the cabin level reaches 20%, the supply is switched off.

Pressure control: When pressure falls below 990 mbar, the nitrogen pump is switched on and a nitrogen/oxygen mixture flows into the cabin until the pressure reaches 1020 mbar. If pressure rises above 1025 mbar, cabin air is vented off until pressure falls below 1020 mbar. By these two means, pressure is maintained in a range of 990 - 1025 mbar.

Carbon dioxide control: A CO_2 level above 1% is highly undesirable for humans and the CO_2 controller operates to keep the CO_2 level below this. Whenever CO_2 rises above 1%, the scrubber is switched on until the CO_2 level is reduced to 0.5%.

Temperature control: When cabin temperature rises above 22° C, the air cooler is switched on. Once on, the cooler stays on until the temperature falls to 20° C, whereupon it is switched off. If the temperature should fall below 19.5° C, the heater is switched on and remains on until the temperature is restored to 21° C. In effect, the heater and the cooler together maintain air temperature in the range 19.5 to 22° C.

Humidity control: The dehumidifier is invoked when relative humidity climbs above 42% and remains in operation until the level is brought down to 38%.

A characteristic of complex systems is that its components do not operate independently in the sense of not affecting each other. Similarly, AMT has been designed to incorporate a number of interactions to make the task environment as realistic as possible. Having a knowledge of these interactions will aid the skilled operator to control the system more efficiently. However, due to the limited time available for performance testing (i.e. 30 min sessions), some artificial changes had to be made concerning the underlying mathematical models used in AMT. The dynamic processes in the system were greatly accelerated and some interactions were magnified in comparison to the natural laws, which represented the basis of the mathematical equations used in the simulation (e.g. Boyles law: increasing pressure causes a rise in temperature).

4.2.1.5 Tasks of the operator

A number of tasks are to be carried out by the operator, which were classified as primary and secondary tasks according to their importance for crew survival. The first two activities listed below were defined as primary tasks, whereas the remaining three are regarded as secondary tasks.

Maintenance of key parameters in target state: The most important task is to maintain the five key parameters within their target state by monitoring their levels and by carrying out interventions if required. It has to be stressed that manual control should only be selected if the operator has observed a disturbance in the normal system operation. Manual control can be very demanding in terms of the mental resources required and should therefore only be carried out if the automatic control system has failed or is not able to deal with the system disturbance effectively.

Fault diagnosis: The second primary task is concerned with diagnostic activities with some significant differences between AMT versions. Essentially, the operator has to diagnose malfunctions of the system by indicating the kind of fault and the component affected.

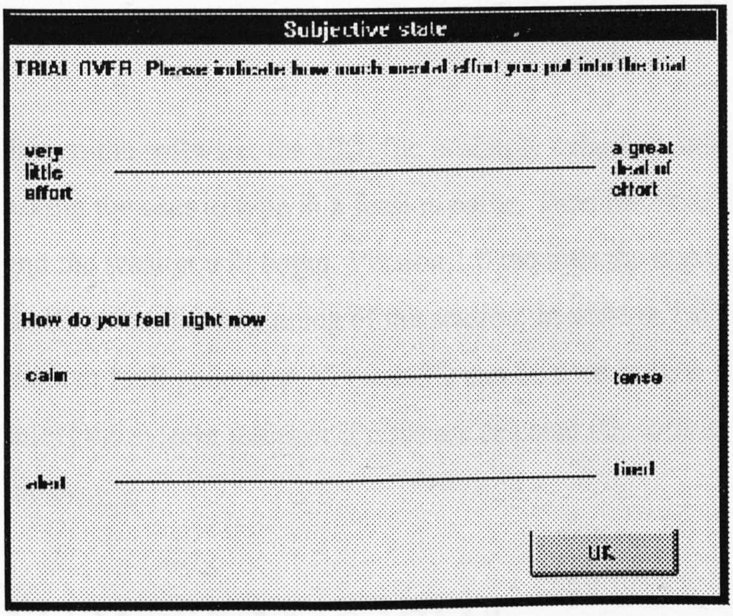
Acknowledgement of alarms: The first secondary task refers to the monitoring of the alarm system. If a warning signal appears on the screen, the operator needs to click upon it immediately. This is to indicate that the alarm was noticed and the parameter concerned will be checked. If it turns out to be a false alarm, no further action has to be taken. If the alarm is genuine, appropriate interventions are to be carried out.

Resource conservation: Oxygen and nitrogen represent a scarce and precious resource, which have to be conserved as far as possible. Any wasteful control strategies are to be avoided since a rapid decline of the reserves during system operation may endanger crew survival.

Tank level recordings / Status reports: This secondary task tests prospective memory by requiring the operator to execute an activity at fixed intervals. It differs slightly between versions with regard to its operational content. In AMT 1.0 the O_2 tank levels are to be recorded at 4-minute intervals, and in AMT 3.0 at 3-minute intervals. In AMT 2.0 reports of the operational state of the system (i.e., whether the system components operate faultfree) are to be made at 2-minute intervals. A prompt at the top of the panel indicates when the next response is due.

4.2.1.6 Subjective state measures

At certain intervals (varies between AMT versions) the operator is asked to complete a short questionnaire to indicate his/her current operational state. It contains three items, which measure the following dimensions of operator state: Mental effort, Anxiety and Fatigue. Responses are to be made on visual analogue scales (100 mm) with the operator simply marking the line with a mouse to indicate his/her current state. The subjective state questionnaire is presented in Figure 4.2.



TRIAL OVER Please indicate how much mental effort you put into the trial

How do you feel right now

Figure 4. 2: Task-embedded subjective state questionnaire

4.2.1.7 Data logging

The data logging facility allows a recreation of the experimental session in its entirety. It automatically records the system state at 10 second intervals (30 sec for version 1.0). The recording comprises the position of the key parameters and the tank levels. Every action taken by the operator (e.g. increasing O_2 inflow, viewing of history display) is also recorded with the exact time of its occurrence. Likewise, if the system carries out an activity (e.g. a warning signal comes on), a record is made in the result file.

4.2.1.8 Operation and technical specification of software

AMT can be run on a PC (386 or more powerful) with a colour monitor. It operates in a windows environment (preferably version 3.1 or later) and requires the instalment of a mouse. The following files are necessary to run it:

for AMT 1.0: AMT1.EXE and VBRUN100.DLL
for AMT 2.0: AMT2.EXE and VBRUN200.DLL
for AMT 3.0: AMT3.EXE and VBRUN300.DLL

The DLL file should be placed into the \windows directory. The EXE file may be placed in any directory. Furthermore, a directory named \AMTDATA needs to be created to store the data files produced after each AMT session. The AMT data files have the format *fname.dat*, with *fname* being the session name entered by the operator / experimenter on the set-up screen.

To run the software, the EXE file is simply called up, which will produce a window, asking the user to type in a session name. Then the 'start' button needs to be clicked upon and the session will begin. Version 1.0 requires the experimenter to set up the system faults before the beginning of the session. In version 2.0 the faults are selected by a random fault generator. Version 3.0 can be used in both ways; the experimenter can select the system faults or if none are selected the random fault generator will set up the faults.

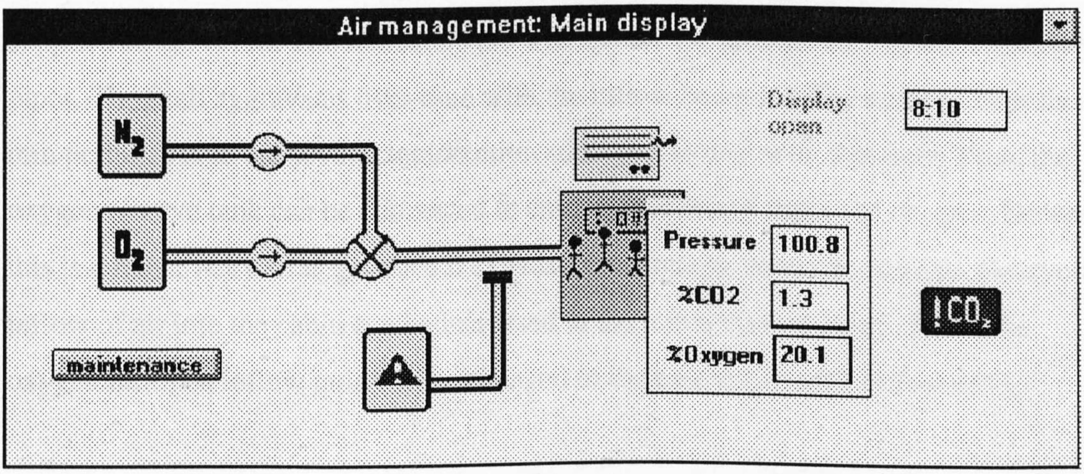
4.2.2 Special features of AMT 1.0

4.2.2.1 Introduction

In this version three key parameters (oxygen, air pressure, carbon dioxide) need to be kept within a pre-defined band. The software allows the experimenter to select from three different system faults on the set-up screen, which will then occur at chosen times during the experimental session. The experimenter can also select between two dialogue control modes (*human-centred* vs *machine-centred*). In the *human-centred* mode all system components are accessible at any time, whereas in the *machine-centred* condition access is severely restricted and controlled by the automatic system. In addition, the set-up screen allows the selection of a number of secondary tasks to vary the task demands of the operator.

4.2.2.2 Operator interface

Main display: The main display consists of the mimic of the life support system, the alarm system, air supply and air conditioning system (see Figure 4.3). It has a number of components, which can be interrogated by clicking on the appropriate icon. The following components, when clicked upon, provide information about the system state: The *cabin display* provides the current values in the space cabin for oxygen, air pressure and carbon dioxide. The *air conditioning system* provides information about the set rate of the air extractor fan and the set range for air pressure in the space cabin. The *mixer valve* indicates the set values for oxygen level in the cabin, the air inflow to the cabin and the inflow ratio of oxygen and nitrogen. The *oxygen tank* and *nitrogen tank* indicate the current tank levels of oxygen and nitrogen. The *oxygen pipe* provides information about the inflow of oxygen into the mixer valve. The *nitrogen pipe* indicates the inflow of nitrogen into the mixer. The *reserve oxygen tank* displays the amount of oxygen left in the tank.



8:10

Pressure 100.8

%CO2 1.3

%Oxygen 20.1

Figure 4. 3: Main display of AMT 1.0

Control panels: The life support system is controlled by means of the following two subsystems: *Air supply* and *Air conditioning*. The air supply system allows the following interventions to be carried out: (1) Setting of oxygen level in cabin for automatic controller, (2) Setting of inflow of air into cabin, (3) Setting of ratio of oxygen:nitrogen inflow into cabin, (4) Use of manual over-ride, and (5) Resetting of system parameters to default values. The air conditioning system allows the following interventions to be carried out: (1) Setting of pump rate for extractor fan, (2) Setting of cabin pressure band for automatic controller and (3) Use of manual over-ride (fan on/off).

4.2.2.3 System failures

AMT 1.0 allows the set up of the following three system failures, which require operator intervention to maintain suitable atmospheric conditions in the space cabin. The exact time of the occurrence of each fault can be specified.

Failure of automatic O_2 controller: The failure of the automatic controller results in the disregard of the set points for O_2, which leads to a decline of the oxygen cabin levels. The operator can detect this by checking oxygen levels in the cabin. When a failure of the automatic controller is identified, the operator has to switch to manual control, which requires a closer monitoring of the system since the oxygen supply has to be switched on and off at appropriate intervals to maintain oxygen levels.

Mixer blockage: The mixer begins malfunctioning by reducing the quantity of air supplied to the cabin. The indicators are a reduced flow through the O_2 and N_2 supply pipes. The fault aggravates over time with the inflow being steadily reduced over the 10 min period. There are several intervention strategies to deal with this fault. First, the oxygen reserve tank can be connected to ensure a sufficient supply of oxygen. However, this is a rather risky strategy since it exhausts the strategic reserves. Second, the overall inflow of air into the cabin can be increased in the mixer control panel. Third, the oxygen:nitrogen inflow ratio can be changed to increase the overall inflow of O_2.

Cabin leak: The leak is located in the air conditioning system with air escaping from the cabin, resulting in a sharp fall of pressure. The first measure is to switch off the air conditioning system to avoid any further escaping of air. This will gradually build up again the air pressure, moving towards its target range. However, at the same time CO_2 is in danger of rising above its target level. This is due to the switching off of the air conditioning system, which would otherwise extract CO_2 from the cabin to maintain levels within a safe range. The operator is therefore required to switch on the air conditioning system from time to time for very brief moments (a few seconds) to reduce CO_2 levels. If the air conditioning system is switched on for too long, cabin pressure will drop too much.

4.2.2.4 Tasks of the operator

In addition to the information given in section 4.2.1.5, the operator tasks specific to AMT 1.0 are described below. This was thought to be useful since AMT 1.0 contains a number of tasks, which have undergone considerable modifications in subsequent versions.

Maintenance of key parameters in target state: The primary task is to keep the three key parameters within pre-defined limits in order to maintain appropriate atmospheric conditions in the space cabin. The critical cabin levels are defined as follows:

Oxygen: 18% (minimum)
Pressure: 96-105% (range)
Carbon dioxide: 3% (maximum)

Acknowledgement of alarms: This refers to the monitoring of system alarms. If a warning signal appears on screen, the operator needs to click on it to indicate that the alarm was noticed and the appropriate parameter will be checked.

Tank level recordings (TLR): Operators are asked to keep a record of the main O_2 and N_2 tank levels at 4-min intervals throughout the session, and to compute oxygen consumption over the previous 4-min period. The recordings need to be carried out exactly when the display-embedded clock strikes the next 4-min time marker. Since there is no further prompt for this task (such as an acoustic signal), this is essentially a prospective memory task.

Oxygen conservation: The consumption of O_2 is measured at the end of the session to provide a measure of effective resource management.

System Failures Log: This task involves the recording of any observed system failures during the session by making a note in a special system failures log book (paper-based). For each identified fault the following information has to be written into the log book: time of occurrence, diagnosis and intervention strategy employed. A column for comments was also provided.

4.2.3 Special features of AMT 2.0

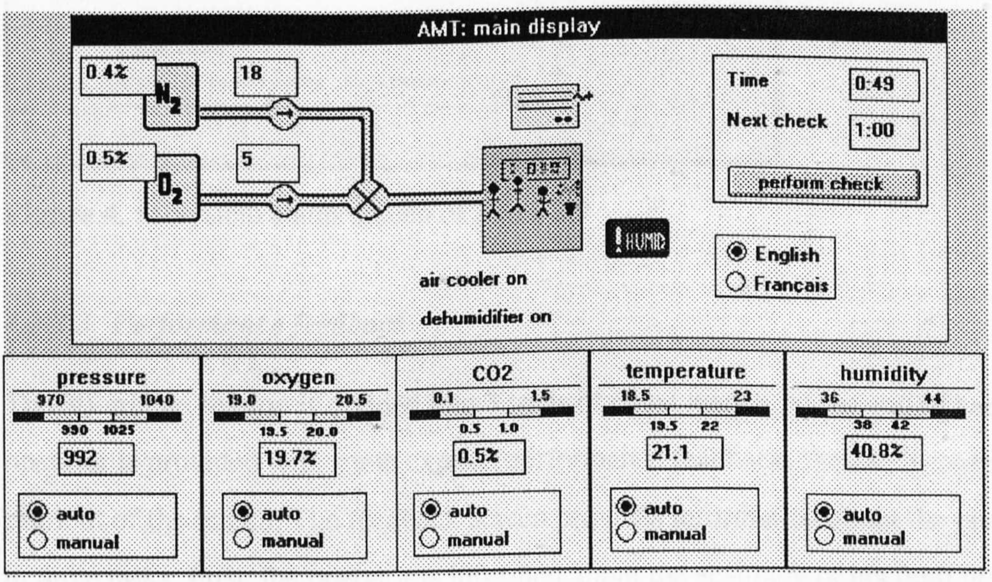
4.2.3.1 Introduction

AMT 2.0 has been developed for the particular needs of the Antarctic over-wintering expedition. The specifications required a version, which can be set up and run easily by the operator him/herself. Furthermore, training time was limited and the project did not allow for the experimenter to administer all the training sessions himself. Therefore, a version with moderate complexity was required, which allowed the operators to acquire sufficient expertise to manage the system adequately within a relatively short period of time.

The session length is 30 minutes, which is divided into three trials of 10 minutes, after which all system parameters are reset to their original value. The software can be run either in English or in French by simply selecting the desired language in the appropriate box on the main display.

4.2.3.2 Operator interface

Main display: The interface of AMT 2.0 consists of the main display, which was already described in section 4.2.1.3, and the control panels of the five key parameters (see Figure 4.4). Resource consumption of O_2 and N_2 is indicated as well as the flow of these gases into the cabin. A panel in the top right hand corner contains the system clock, a prompt for system checks and the button for invoking the window to complete the check.



0.4%
18
0.5%
5

Time 0:49
Next check 1:00

● English
○ Français

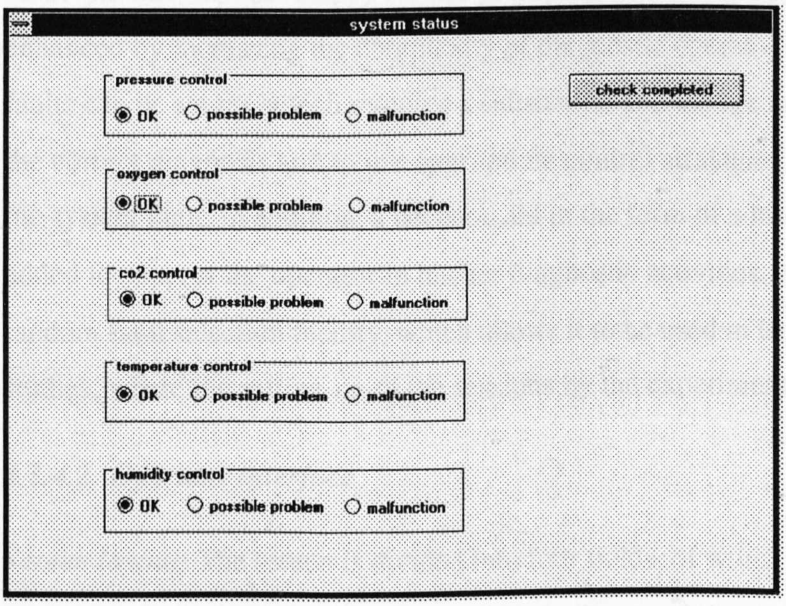
air cooler on
dehumidifier on

pressure: 992 ● auto ○ manual
oxygen: 19.7% ● auto ○ manual
CO2: 0.5% ● auto ○ manual
temperature: 21.1 ● auto ○ manual
humidity: 40.8% ● auto ○ manual

Figure 4. 4: Main display of AMT 2.0

Control panel: The control panel of each parameter (see Figure 4.4) consists of the graphical indicator of the different safety zones and the current levels (see section 4.2.1.2). Furthermore, it contains a dialogue box to select between automatic and manual control mode. The selection of manual control will invoke a further window, allowing the increase or reduction of parameter levels, during which the automatic control system is over-riden (see pressure control panel in Figure 4.4). For three of the control systems (O_2, CO_2 and humidity), interventions can only be carried out in one direction (e.g. increase of oxygen) since the corresponding parameter tends to move only one way (i.e. O_2 steadily declines without intervention from the system).

System status screen: The screen is invoked by clicking on the 'perform check' button on the main display. The operator is asked to give an assessment of the operational state of each of the five subsystem by indicating any possible malfunctions (see Figure 4.5).



pressure control
● OK ○ possible problem ○ malfunction

oxygen control
● OK ○ possible problem ○ malfunction

co2 control
● OK ○ possible problem ○ malfunction

temperature control
● OK ○ possible problem ○ malfunction

humidity control
● OK ○ possible problem ○ malfunction

Figure 4. 5: System status screen of AMT 2.0

4.2.3.3 Performance feedback

The 30-minute session is divided into 3 trials of equal length. At the end of each trial the operator is provided with some explicit performance feedback. Picture 5 contains a printout of the screen. The feedback screen provides information about the performance in three tasks. First, the operator is informed about the accuracy of his fault diagnosis. Second, a score (in %) is provided for the time out of range for each of the five parameters. Third, scores for resource consumption are given, which also includes a figure for energy usage (not measured in AMT 3.0).

4.2.3.4 System failures

All system failures occurring in this version are caused by a malfunctioning automatic controller, resulting in set points being ignored and the corresponding parameters moving out of range. Within each experimental session (= 30 min) the software is programmed to select two trials (1 trial = 10 min) with two automatic controllers failing, and one with four malfunctioning automatic controllers. Once an automatic controller has failed, it remains in this state until the end of the trial.

4.2.4 Special features of AMT 3.0

4.2.4.1 Introduction

AMT 3.0 has been developed for use in Experiment 2. The version has been considerably enhanced by increasing the complexity of diagnostic operator tasks. A total of 18 system faults can be set up, which requires a substantial knowledge from the operator to make the correct diagnosis and to use appropriate control strategies to cope with the impact of the system fault. In addition, a decision aid in the form of a history display has been added to support the operator in his/her diagnostic activities. AMT 3.0 also contains a random fault selection facility, which allows it to be used in isolation and confinement studies, where the setting-up of the sessions by the experimenter is not a feasible option.

4.2.4.2 Operator interface

Screen layout: The screen is divided into four fields, of which each has been allocated to a main functional display (see Figure 4.1). In the top screen area, the main display is located on the left and the screen manager on the right hand side. In the bottom area of the screen, the history display is situated to the left and the control panels are to be found in the right hand corner.

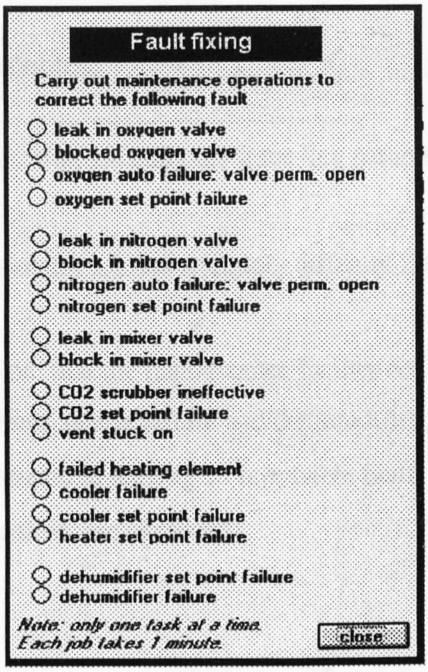
Main display: The design of the display is very similar to version 2.0 with three additional features (see Figure 4.1). In the bottom right corner, a prompt and a dialogue box is provided to carry out the O_2 tank level recordings. Below the mimic there is the 'flow rates' button, which, when clicked upon, displays the current readings of the flow meters and the current tank levels for a duration of 10 seconds. The function of the 'maintenance' button is described below.

Screen manager: The number of displays available for interrogation and intervention have considerably increased compared to previous versions. Therefore, in order to avoid a cluttered display, the screen manager facility was introduced to control the number of displays present on the screen. Clicking on the appropriate button invokes the requested display, which will remain on screen for 30 seconds. The main display, however, is permanently present. Limiting the time period during which a display is visible has the advantage of being able to record the information sampling behaviour of the operator.

Control panels: Only one control panel is displayed at a time for reasons of limited screen space. As in AMT 2.0, the control panel of each parameter (see Figure 4.1) consists of the graphical safety zone indicator and the current levels. The dialogue box facilities underneath allow the operator to switch to manual control (as in AMT 2.0) but also to select between four different settings (high, medium, low and zero). This enables the operator to make a more finely adjusted response to system disturbances. Again, three out of the five control systems (O_2, CO_2 and humidity) operate unidirectionally.

History display: The history display (see Figure 4.1) provides graphical information about the development of each key parameter over the previous four minutes. It contains lines to indicate the different safety zones (red, yellow and white). The history display is mainly used for two purposes. First, it gives an overview of the current state of the main parameters. Second, it provides historical information, which can aid in the diagnostic process of fault finding.

Maintenance function: This facility allows the operator to repair system faults, once a particular fault has been identified. Each repair takes one minute to complete, during which no other repair can be carried out. The maintenance function is able to rectify 19 different system faults (see Figure 4.6).



Carry out maintenance operations to correct the following fault

○ leak in oxygen valve
○ blocked oxygen valve
○ oxygen auto failure: valve perm. open
○ oxygen set point failure

○ leak in nitrogen valve
○ block in nitrogen valve
○ nitrogen auto failure: valve perm. open
○ nitrogen set point failure

○ leak in mixer valve
○ block in mixer valve

○ CO2 scrubber ineffective
○ CO2 set point failure
○ vent stuck on

○ failed heating element
○ cooler failure
○ cooler set point failure
○ heater set point failure

○ dehumidifier set point failure
○ dehumidifier failure

Note: only one task at a time.
Each job takes 1 minute.

Figure 4. 6: Maintenance facility of AMT 3.0

4.2.4.3 Performance feedback

The operator receives some explicit performance feedback at the end of each session. It contains the following scores: (1) Percentage of time out of range for the five main parameters together, (2) consumption of resources (O_2 and N_2), (3) percentage of annunciators (warning signals) acknowledged, and (4) Percentage of O_2 tank level recordings carried out on time.

4.2.4.4 System failures

The software allows the setting up of 18 different faults[3], of which a complete list can be found in Appendix 1. The faults may be grouped into five categories according to the underlying technical problem: Set point failures, leaks, blocks, closure failures and control panel failures. A set point failure essentially represents a failure of the automatic controller, which require manual control action to keep the parameter in range. A leak refers to the escape of gas at one point in the pipework. If the flow of gas through the pipes is reduced when passing a valve, a block has occurred. The opposite defect to a block is the closure failure, where a valve cannot be entirely closed. The last group of failures refers to a complete breakdown of the control panel, which means that this parameter cannot be directly controlled any more. In addition, a failure of the maintenance facility can be programmed in so that the operator has to deal with the fault for an extended period of time without being able to repair it.

Endnotes:

---

[1] Version 1.0 had only three key parameters (O_2, pressure and CO_2).

[2] Version 1.0 had a slightly different layout of the control panels (see section 4.2.2 for details).

[3] Note that the maintenance facility contains 19 faults. Three of these faults (heater set point failure, heater failure and dehumidifier failure) do not occur in the system and represent distractors, leaving 16 rectifiable faults (i.e. the two control panel failures, which cannot be repaired, are excluded).

Chapter 5:

Sleep Deprivation and Dialogue Control

5.0 Summary

Since partial loss of sleep is a common occurrence in extended spaceflight, an experiment was conducted to examine the effects of sleep deprivation on performance in the AMT environment. The design of the human-machine interface may have a moderating effect on performance in that high dialogue control may offset any possible negative effects of sleep deprivation. Two kinds of interface were used. The first one was called *operator-centred (O-C)* and permitted a high level of dialogue control. The second one, with a low level of dialogue control, was labelled *machine-centred (M-C)*. Using an O-C interface would mean higher involvement for the operator in the control process, allowing the maintenance of a more up-to-date mental model of the current system state, which may prove advantageous in the event of system failures. In addition, task difficulty (defined as the difficulty of a fault state) was manipulated as a third variable to be able to determine any possible interaction effects with interface design and sleep deprivation. A repeated measurement design was used to test 16 participants under all experimental conditions. Each testing session lasted for 2 hours.

Summarising the results and interpreting them in the context of the model of compensatory control provided a very complex picture of adaptive processes, which were used by the operator to cope with the varying task demands. The results suggested that effort expenditure and adaptive changes in control and information sampling strategies were primarily used to cope with the sub-optimal working conditions caused by M-C and sleep loss. Only a few secondary task decrements were observed, indicating that operators largely attempted to maintain performance on all tasks rather than adopting a selective strategy to focus resources on high-priority tasks. The picture was in line with the general predictions of the theoretical framework but a number of results found were in contrast to some of the more specific predictions. The complex picture found also advocates the use of a complex methodology (as used in this study) rather than adopting a simpler approach (e.g. only looking at performance), since the results of the latter may not reflect the complexity of human behaviour in simulated as well as real work environments.

5.1 Introduction

5.1.1 Sleep loss and performance

Sleep quality has been a major concern throughout the history of manned spaceflight because of the many factors (such as noise, altered sleep-wake patterns, lack of zeitgebers and micro-gravity) which may impinge on the quality of sleep. Crew members reported reductions in sleep quantity as well as quality on a number of missions (Connors et al 1985; Bluth and Helppie 1987). This means an accumulation of partial sleep deprivation over a prolonged period of time with an increasing risk of performance decrements. On some Soviet missions, the work-rest schedules had to be changed temporarily by increasing the time allowance for sleep to 12 hours, and sleeping pills were taken by the cosmonauts to reduce the amount of lost sleep (Bluth and Helppie 1987). However, it also emerged that sleeping patterns varied considerably between missions and crew members.

The effects of sleep loss are manifold, resulting in mood changes, subjective feeling of sleepiness, performance impairment, etc. (Carskadon and Roth 1991; Pivik 1991). The main question for this thesis is how performance is affected by loss of sleep. In a review of the literature on the effects of sleep loss on performance, Dinges and Kribbs (1991) draw the general conclusion that sleep deprivation can impinge on virtually all cognitive tasks, even if they are of very short duration. The type of cognitive tasks used in these studies included mental arithmetic (Dinges, Whitehouse, Orne and Orne 1988), logical reasoning (Angus and Heslegrave 1985), reaction time (Williams, Lubin and Goodnow 1959) and tracking (Mullaney 1983). However, Hockey (1986) argued that it has been quite difficult to demonstrate an appreciable effect of sleep loss on performance (at least on short duration tasks) while the probability of a performance decrement rises with increasing time on task. There is little argument about the claim that the impact of sleep loss becomes more pronounced in long duration tasks (Wilkinson 1992). Wilkinson therefore demanded a task duration of at least 30-60 min to be sufficiently sensitive to sleep loss. Other literature reviews also came out in support of a time-on-task decrement as a result of sleep loss (Hartley, Morrison and Arnold 1989; Dinges 1992).

There has also been a debate about motivation as a critical factor of performance with some researchers arguing that increasing motivation (e.g. through incentives, comparative feedback) could lead to considerable performance improvements even under conditions of sleep deprivation (Horne 1991, Gaillard and Steyvers 1989). However, this positive effect may disappear with increasing duration of the task (Dinges and Kribbs 1991), indicating the limited usefulness of these measures for extended spaceflight. It has also been shown in a number of sleep deprivation studies that paced tasks (i.e. where the participant has no control over the pace) resulted in a higher number of errors than unpaced tasks because participants were not able to slow down the speed of the task to maintain accuracy levels (Williams et al 1959). This suggests some adaptive strategy (slowing down) used by humans to combat the possible negative effect of sleep loss on accuracy. The human adaptive regulatory systems appeared to have been working successfully in emergency situations, where humans were observed to be able to postpone sleep for quite long periods of time (1-2 days) without showing overt evidence of performance decrements (Dinges and Kribbs 1991). This was attributed to increased effort expenditure.

Most lab-based studies used total sleep deprivation although accumulated partial loss of sleep is more common in spaceflight. This is not regarded as a major drawback since there has been evidence that accumulated partial sleep deprivation over some days showed very similar effects to total sleep deprivation of the same amount (Connors et al 1985). However, others have argued that sleep continuity rather than sleep duration is the critical factor for performance (Campbell 1992), suggesting that partial sleep deprivation is of a lesser concern than total sleep loss. Most studies found in the research literature have used fairly simple tasks (vigilance, reaction time, short-term memory, etc.) and rarely employed more complex task environments as planned in this thesis. This makes it more difficult to predict the effects of sleep deprivation on human behaviour in multiple-task environments.

5.1.2 Dialogue control and the operator interface

In a spacecraft, a number of automated systems are in operation to ensure crew survival (see Chapter 1 for a more detailed discussion). One can expect that more systems of this kind and with higher complexity levels will be used on future missions. This is largely because reliance on earth supplies needs to be reduced (in particular for long-term missions and distant space stations), which requires the recycling and production of various resources (European Space Agency 1994; Cheston and Winter 1980). This would require the design of complex process control systems. In addition, the complexity of currently existing systems (life support, thermal control) will increase since the size of space stations is expanding to accommodate larger crews with their increased need for more personal space during extended missions. Following from the general discussion of interface design in Chapter 2, this raises more specific questions about the appropriate level of dialogue control in future space systems.

A high level of control is generally considered as desirable in the research literature. It is argued that control at work is associated with well-being and improved performance (Karasek 1979, Frese 1989a). In particular, increased level of control (or decision latitude in Karasek's terms) were found to offset the strain symptoms associated with high workload (Karasek 1979). Control is a multi-faceted concept, covering aspects ranging from the freedom to select a particular intervention strategy to issues of work organisation in an organisational context (Frese 1989b). Focusing on the here relevant issue of dialogue control, there are two main benefits of high dialogue control for the human operator. First, staying in-the-loop allows the operator to maintain an up-to-date mental model of the current system state, a prerequisite for good control and diagnostic performance (Wirstad 1988). Second, it allows the operator a more active role in scheduling work tasks, avoiding problems associated with peak demands as well as with underload by anticipating or delaying certain tasks (see also Spérandio 1971, 1978). However, in contrast to the generally accepted desirability of enhanced dialogue control, real systems have often not had these features. This was because it had been argued (predominantly in the engineering community) that the reduction of operator involvement would decrease the potential for human error (see Reason 1990) and operators would have less opportunity to cause disturbances of system states by using poor control strategies at times when no intervention was required (see Hockey and Maule 1995). This has led to system designs where operator intervention is only possible when system states are disturbed and the machine is not able to solve the problem. During all other times, the operator is in the position of a system monitor.

Overall, the evidence points clearly in~~to~~ the direction of enhanced dialogue control being of great benefit but one also needs to be aware of its inherent drawbacks. One may therefore consider the task of finding the appropriate level of operator control in a highly-automated system as one of the most difficult ones faced by system designers, given its considerable implications for overall human-machine system performance.

Closely related to dialogue control is the presentation and integration of information to achieve an optimal level of system informativeness. Dialogue control also covers access to this information, and restricted access by the interface (because the system designer thought that this information was not relevant) has implications for keeping an up-to-date mental model of the system state. Therefore, restricted dialogue control often means reduced system feedback. These design issues require careful consideration since they often involve trade-offs between conflicting goals (e.g., increased transparency may lead to information overload with the operator unable to find the significant data; Woods 1986).

The reader is also referred to Chapter 2, where the issue of designing highly-automated work environments and its implications have been discussed in a wider context. In the domain of spaceflight, the issue of dialogue control gains additional importance when one considers that astronauts represent a group of highly-qualified and highly-trained professionals, whose levels of qualification exceed by far those of the average process control operator. Therefore, the need for crew members to be involved in system operation is to be considered of even greater significance in spaceflight than in some analogue work environments.

In view of the evidence from the literature, we would not expect any effects of sleep loss on the primary tasks but rather some performance impairment on secondary tasks. It was also hypothesised that sleep deprivation would lead to some strategic changes in control and information sampling behaviour. A human-machine interface with high dialogue control was expected to have a moderating effect in that it may offset any negative impact of sleep deprivation on performance. A main effect of interface on primary task performance was also predicted since enhanced dialogue control would allow the operator to adopt a more active role in task management.

5.2 Method

5.2.1 Design and data analysis

A repeated measurement design was used with three independent, within-subject variables: *Sleep deprivation, interface design* and *fault type*. Sleep deprivation was varied at two levels (normal sleep vs one night sleep deprivation), dialogue control at two levels (operator-centred interface vs machine-centred interface) and fault type at four levels. The level of fault type corresponded to a particular fault state in the simulated life support system (faultfree, autofailure, mixer block, leak).

AMT 1.0 was used as the experimental task in this study. Participants were tested four times (each session lasted 2 hours) with sleep deprivation and interface being varied between sessions to allow for all possible combinations of the two variables. Testing took place in groups of 4 (crew) with each crew going through the four experimental conditions in a different order to counterbalance any possible order effects (e.g. caused by practice). Table 5.1 contains the experimental plan with the order of the different testing conditions.

Crew	Session			
	1	2	3	4
A	O-C/NS	O-C/SD	M-C/SD	M-C/NS
B	M-C/SD	O-C/NS	M-C/NS	O-C/SD
C	O-C/SD	M-C/NS	O-C/NS	M-C/SD
D	M-C/NS	M-C/SD	O-C/SD	O-C/NS

Table 5.1: Order of experimental sessions for the 4 crews (O-C = Operator-centred; M-C = Machine-centred; NS = Normal Sleep; SD = Sleep Deprivation)

The variable 'fault type' was varied within sessions. Each 2-hour session consisted of three 40-min periods, which represented the basic unit for the experimental set-up plan. The duration of fault states within each 40 min period was as follows: Faultfree (20 min), O_2 control system failure (10 min) and mixer block or cabin leak (10 min).

The onset of a particular fault period followed an experimental plan, which was unknown to the participants and therefore they were unable to anticipate the beginning and the type of a particular fault state. The fault phases always lasted for an uninterrupted period of 10 minutes, whereas the faultfree phase was divided into three subphases adding up to

20 minutes. This is an example of the set-up of a 40-min subperiod: Faultfree (min 0-5), O_2 control system failure (min 5-15), faultfree (min 15-27), cabin leak (min 27-37), faultfree (min 37-40). In each 40-min subperiod either a mixer block (33.3%) or a cabin leak (66.7%) occurred. Consequently, over a full experimental session (3x40 min) each participant had to deal with one mixer fault and two cabin leaks.

Although analysis of variance is very robust against violations of the assumptions of normality and homogeneity of variance (Howell 1992), some of the data needed to be transformed by a square-root or logarithmic transformation to achieve a reduction of the skewness of the distribution and to reduce heterogeneity of variance. This was because many of the DVs had a positively skewed distribution. This is very common with performance measures, indicating deviations from an optimal performance. The positive side of the distribution is limited to zero (optimal performance), whereas the negative side is open to very large scores (poor performance). However, for reasons of clarity tables and graphs always contain untransformed data.

5.2.2 Participants

Participants were mainly recruited among the Hull University postgraduate student population. Some basic selection principles were applied to ensure that participants matched elementary criteria for the selection of trainee astronauts. Therefore, they were required to be computer-literate as well as to have some basic knowledge of the functioning of process control systems. Furthermore, the participants should be studying or have completed their studies in a pure or applied science. They were all invited to an informal interview to ensure that they fulfilled the selection criteria and to give them some more detailed information about the experiment.

As a result of the selection procedure, a total number of 20 participants were recruited of which 16 (14 male, 2 female) successfully completed training and took part in the experiment. Their ages ranged from 23 to 34 with a mean of 25.1 years. Fourteen were pursuing a higher degree, whereas the other two worked as research assistants or post-docs. The breakdown of their academic background was as follows: Engineering (5), computer science (4), chemistry (3), biology (2) and physics (2).

5.2.3 Training

Training was given to the participants in small groups of two or three in sessions of approximately one hour. Each participant came to the laboratory for six training sessions, which corresponded to a total training time of about six hours. Participants had their own PC available during the training sessions to give them as much hands-on practice as possible. Training aimed to minimise theoretical explanations that were not accompanied by practice on the task.

In the first training session the experimenter explained the essential features and functions of the software before the participants were asked to operate the system in a faultfree mode. The participants were encouraged to change the settings of the system to improve their understanding of the relationship between the system components. They were asked to switch off the automatic controller and operate the system manually to prepare them for future system failures.

In the following three training sessions they were confronted first with simple faults, before they moved on to the more difficult ones. Appropriate intervention strategies were discussed and participants were given further help if they did not manage to employ a suitable strategy. After having gained some familiarity with the different fault states, they were introduced to the secondary tasks. Whereas some of the secondary tasks were of a very simple nature (e.g., acknowledgement of alarms), others required considerable training time (e.g., system failures log) to reach satisfactory performance levels.

The last two training sessions were devoted to the integration of all the embedded tasks under the different fault conditions and the introduction of the 'machine-centred' mode of operation. This provided the participants with sufficient practice on both control modes of AMT and also gave them sufficient time to practise the different tasks.

5.2.4 Independent variables

Sleep deprivation

The sleep deprivation condition was operationalised as one night of total sleep loss, preceding the experimental session. The participants were kept awake in a university-based laboratory. During the supervised vigil, the participants were free to structure their time within the constraints of the experiment. They typically engaged in activities such as reading, watching TV and videos, playing games and doing jigsaws. Participants were provided with tea, coffee and light snacks. However, the consumption of alcohol was not allowed. The schedule for the sleep deprivation condition was as follows:

2200-2230: Arrival at laboratory
2230-0800: Supervised vigil
0800-0830: Breakfast in laboratory
0830-0900: Briefing, preparation for work session
0900-1200: Experimental session
1200-1230: Debriefing

After a sleep deprived night, participants were not tested for at least four days, providing enough recovery time to avoid any carry-over effects.

Interface design

In the M-C interface, operator access to control panels and information sources was restricted by the automatic system. The operator could only intervene during abnormal system states, which was indicated by the appearance of an annunciator. In addition, participants were deprived of certain system information (such as levels of key parameters, settings of the automatic controller) during normal system operation. Again, they were allowed only access to these data during abnormal system states. In the O-C interface condition, however, participants could interrogate all system components at any time, and all system control panels were permanently accessible.

Fault type

The following four different fault levels occurred in the experiment: (1) *Faultfree:* The automatic controller operated perfectly well and no system component failed or was defective. (2) *O_2 set point failure:* The automatic controller of the O_2 supply failed, requiring the operator to use manual control to maintain adequate oxygen levels. (3) *Cabin leak:* A leak in the cabin causes air to escape, resulting in a sharp fall in cabin pressure. (4): *Mixer blockage:* A partial blockage (which is deteriorating) of the mixer reduces air supply to the cabin.

5.2.5 Dependent variables

As outlined in Chapter 4, AMT allowed for the selection of a large number of variables for the analysis. In correspondence with the study rationale, dependent measures from the following domains were covered in the analysis: *Primary task performance, secondary task performance, system control and information sampling behaviour* and *subjective operator state.* The dependent variables are now briefly described.

Primary task performance. (1) *Parameter control failures* refer to the percentage of time each of the three key parameters has moved out of its target state (i.e. is in the red zone).

Secondary task performance. (1) *Annunciator acknowledgement* is a reaction time measure to indicate the time needed to respond to an alarm signal. (2) *Tank level recording* is a prospective memory task, where recordings were scored for accuracy of timing by using the following categories: Early (-5 sec or more), correct (+/- 5 sec), late (+5 sec to +20 sec) or omissions (more than 20 sec late or missed). (3) *Oxygen consumption* refers to the amount of oxygen consumed during system operation. (4) *System failure log* is a measure of diagnostic accuracy in system fault identification.

System interrogations and interventions. (1) *Control actions* refer to the frequency of system interventions. (2) *Cabin viewing* measures the frequency of interrogations of the cabin status (levels of key parameters).

Subjective state. The following three subjective state measures were collected after each working session: (1) *Anxiety,* (2) *Fatigue,* and (3) *Mental effort.*

5.3 Results

5.3.1 Primary task performance

Parameter control failures (PCF)

Maintaining the three cabin parameters (O_2, Pressure and CO_2) in a predefined target state was defined as the primary task. The overall picture was that O_2 (8.4%) and CO_2 (4.5%) showed a considerably smaller number of PCFs compared to pressure (17.3%), reflecting the difficulty of the participants to control cabin pressure. This was due to a number of factors. First, pressure was a more volatile variable during the occurrence of faults directly affecting it (such as the cabin leak), resulting in very sharp drops, which required highly-skilled control behaviour to avoid PCFs. Second, unlike the two other parameters pressure had to be controlled within a bipolar range. A distinction of PCFs into lower and upper deviations showed an approximately equal distribution of control errors. Third, pressure was affected by both subsystem (air supply and air purification), whereas O_2 and CO_2 were only affected by one of them.

The percentage of PCFs for the three parameters was averaged to obtain an overall PCF index and subsequently analysed by analysis of variance. After a square-root transformation, the results showed that there was no difference between the PCF values across the main testing conditions. Slightly more control errors were committed under O-C (13.5%) than M-C (12.1%) but the difference was not significant (F=3.73; df=1,15; p>.05). Sleep deprived participants showed a marginally poorer control performance (13.2%) than when sleep was normal (12.5%), but again the difference was not significant (F<1). There was no interaction between sleep deprivation and interface (F<1). However, as can be seen in Figure 5.1, a strong effect of fault type was observed (F=23.5; df=3,45; p<.001) but no interaction with interface or sleep. Post-hoc LSD-tests revealed that all group means were significantly different from each other (p<.05), except for the difference between the fault-free condition and O_2 set point failure.

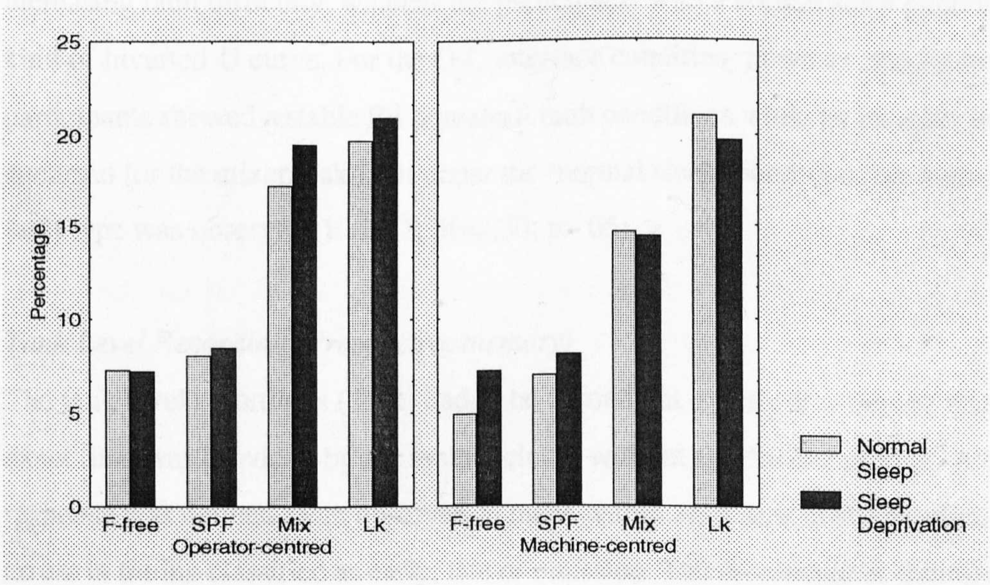


Figure 5.1: Mean PCF (%) as a function of interface, sleep and fault type (F-free = fault-free, SPF = O_2 set point failure, Mix = Mixer blockage, Lk = Cabin Leak)

A separate analysis of each of the three parameters was carried out to determine whether control of these was affected differently by the independent variables. The analysis showed clearly that this was not the case. They indicated the same pattern as the composite measure described above with a strong effect of fault type and the absence of any effects of sleep deprivation and interface.

5.3.2 Secondary task performance

Annunciator acknowledgement (reaction time)

The data were subject to a logarithmic transformation to reduce variance heterogeneity. In order to stabilise the data for the less frequently occurring fault states (mixer blockage and cabin leak), the data of these two variables were averaged into one variable called *plant fault.* The analysis revealed a strong effect of interface (F=19.75; df=1,15; p<.001), with participants taking considerable longer to react to warnings under the M-C interface (14.5 sec) than under the O-C interface (7.0 sec). No main effect of sleep deprivation was found (F<1) but there was an interaction between interface and sleep (F=7.56; df=1,15; p<.05). It occurred because sleep deprived participants working with an M-C interface needed about twice as long to respond than when tested in the three other conditions. Furthermore, a complex three-way interaction between the main independent variables was observed (see Figure 5.2): F(2,30)=8.27; p<.005. It showed that under M-C the sleep deprived participants displayed the expected pattern of increased RT with increasing fault difficulty, whereas the participants with a normal sleep pattern showed a kind of inverted-U curve. For the O-C interface condition, however, the sleep deprived participants showed a stable RT across all fault conditions while an increase in RT occurred for the mixer/leak fault under the 'normal sleep' condition. No main effect of fault type was observed (F=2.31; df=2,30; p>.05).

Tank Level Recording (Prospective memory)

The tank level recordings (TLR) had to be carried out exactly at 4-minute intervals (the exact time was provided by the system clock) without any further prompt. A correct response was defined as an observation made within +/- 5 secs of the required time, with errors in timing classified as early, late or omission. The percentage of timing errors is presented in Table 5.2. It shows that only about a third of the recordings were made on time. A considerable number of TLR were made early or late, with only few completely omitted.

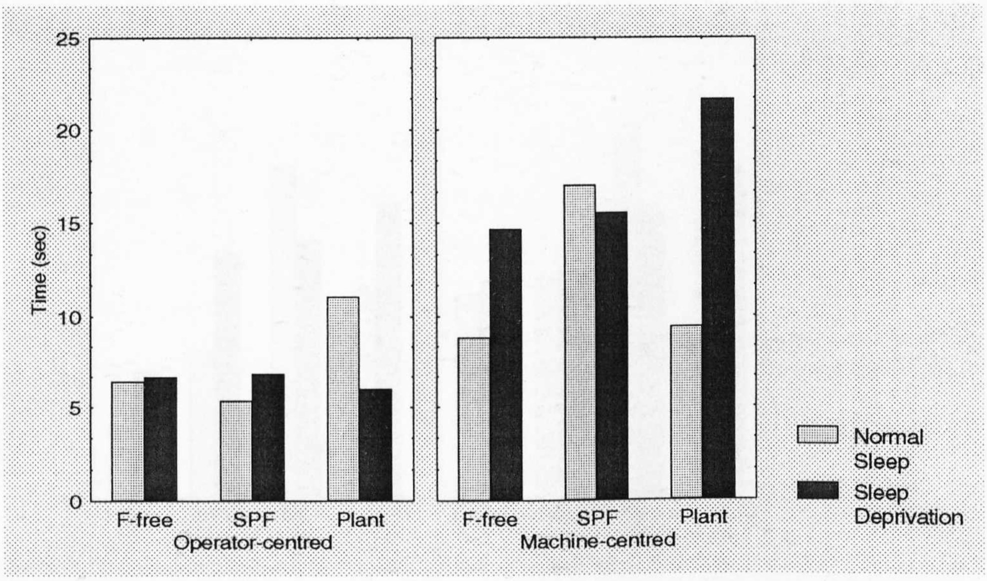


Figure 5.2: Mean RT (sec) as a function of interface, sleep and fault type (F-free = fault-free, SPF = O_2 set point failure, Plant = Plant fault)

	Operator-centred		Machine-centred	
	Normal Sleep	Sleep Deprived	Normal Sleep	Sleep Deprived
Correct	39.6	33.3	29.9	28.7
Early	32.6	39.9	41.2	50.9
Late	22.4	20.1	24.6	17.3
Omission	5.3	6.6	4.1	3.1

Table 5.2: Accuracy of tank level recordings (%) as a function of interface and sleep

Applying the strict criterion of accurate recordings (+/-5 sec), a three-way ANOVA revealed that somewhat more incorrect TLRs were made under M-C control (73.2%) than under O-C control (68.2%). However, the difference was not significant (F=2.24; df=1,15; p>.05). No difference between sleep conditions was found (F<1). Figure 5.3 shows that inaccurate recordings increased with fault difficulty across all conditions. Sleep deprived participants committed the largest number of prospective memory errors during the cabin leak, whereas most errors were made during the mixer blockage under the normal sleep condition. A main effect of fault type proved to be significant (F=6.41; df=3,45; p<.001), planned comparisons indicating a linear trend (F=36.6; df=1,15; p<.001). No interactions were found. A further analysis using a less strict criterion (no of omissions) showed the same pattern of results with only a main effect of fault type being significant.

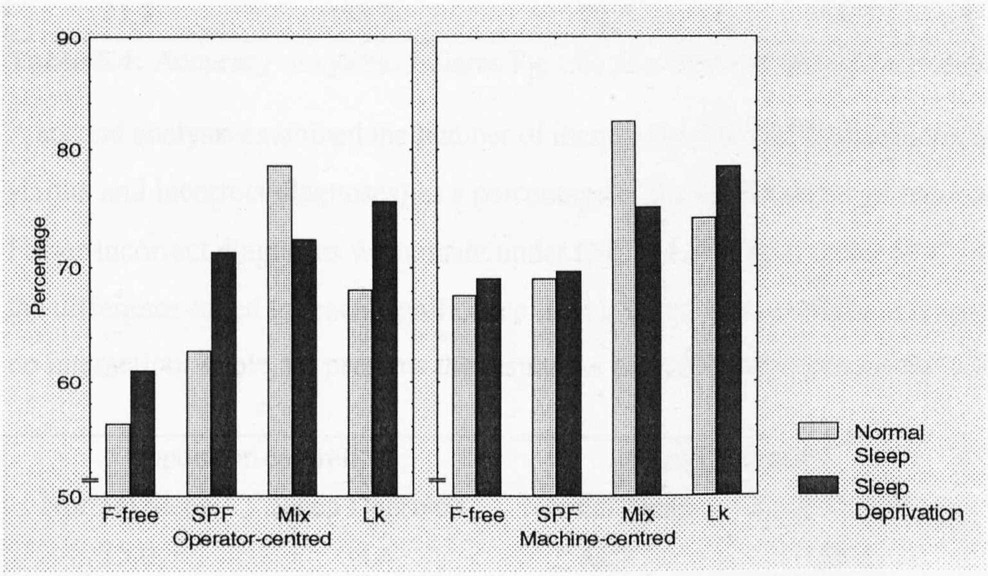


Figure 5.3: Incorrect tank level recordings (%) as a function of interface, sleep and fault type (F-free = fault-free, SPF = O_2 set point failure, Mix = Mixer blockage, Lk = Cabin Leak)

Oxygen Consumption

As shown in Table 5.3, O_2 consumption varied little between conditions. No main effects or interactions were found (all F<1).

Operator-centred		Machine-centred	
Normal Sleep	Sleep Deprived	Normal Sleep	Sleep Deprived
66.7	66.3	67.0	67.3

Table 5.3: Oxygen consumption during work sessions (%) as a function of interface and sleep

System Failure Log (Fault identification)

The system failure logs were analysed by calculating the number of correctly identified system faults. A two-way analysis of variance revealed that significantly more faults were identified under the most favourable condition (O-C & normal sleep), compared to the three other conditions. This resulted in a significant interaction effect (F=7.48; df=1,15; p<.05). No main effects were observed neither for control (F<1) nor for sleep (F=1.39; df=1,15; p>.05). The results are presented in Table 5.4.

Operator-centred		Machine-centred	
Normal Sleep	Sleep Deprived	Normal Sleep	Sleep Deprived
71.8	58.7	62.5	64.3

Table 5.4: Accuracy of system failures log (%) as a function of interface and sleep

A second analysis examined the number of incorrectly reported faults (including false alarms and incorrect diagnoses) as a percentage of the total number of recorded faults. Fewer incorrect diagnoses were made under O-C (34.7%) than under M-C (40.2%) but the difference failed to reach significance (F<1). There was no effect of sleep (F<1) and no interaction. Table 5.5 presents the results for the different experimental conditions.

Operator-centred		Machine-centred	
Normal Sleep	Sleep Deprived	Normal Sleep	Sleep Deprived
35.0	34.4	41.9	38.4

Table 5.5: Inaccurate reports of fault occurrences (%) as a function of interface and sleep

5.3.3 System interrogations and interventions

Control actions

A three-way ANOVA revealed that more control actions were carried out under sleep deprivation than under normal sleep. Although the difference was small (0.38 vs 0.34 CA/min), it reached statistical significance (F= 5.62; df=1,15; p<.05). Contrary to expectations, no significant difference was found for interface (F=1.65; df=1,15; p>.05) although marginally more control activity was observed under O-C. As can be seen in Figure 5.4, the effects of sleep deprivation were more pronounced under M-C but the interaction was not significant (F=2.70; df=1,15; p>.05). A main effect of fault type was found with control activities during fault conditions increasing in the following order: faultfree (0.21 CA/min), mixer block (0.34 CA/min), O_2spf (0.43 CA/min) and leak (0.48 CA/min). This was highly significant (F= 43.75; df=3,45; p<.001) and post-hoc LSD-tests showed that each group mean was different from the other at the p<.05 level. No interactions were found.

A separate analysis may be carried out for the two subsystems of the life support system (*air conditioning* and *air supply*). This showed that there was no effect of sleep deprivation for air supply but air conditioning CAs increased under sleep deprivation by ca. 20% (F= 10.9; df=1,15; p<.005). A further distinction may be made between two control modes: *changing the settings of the system* (e.g. increasing the O_2 inflow from 100 to 250) and *on/off control actions*. The former control mode is considered more demanding since it is more similar to feedforward control, requiring a better mental model to predict the impact of different parameter settings. Operators generally preferred the on/off control actions (0.47 CA/min) to setting changes (0.25 CA/min). Whereas the frequency of change setting CAs remained stable across the two conditions of the variable sleep, the number of the less demanding on/off CAs increased under sleep deprivation (0.44 to 0.51). Although the size of the increase was rather modest, it proved to be significant (F=10.4; df=1,15; p<.01).

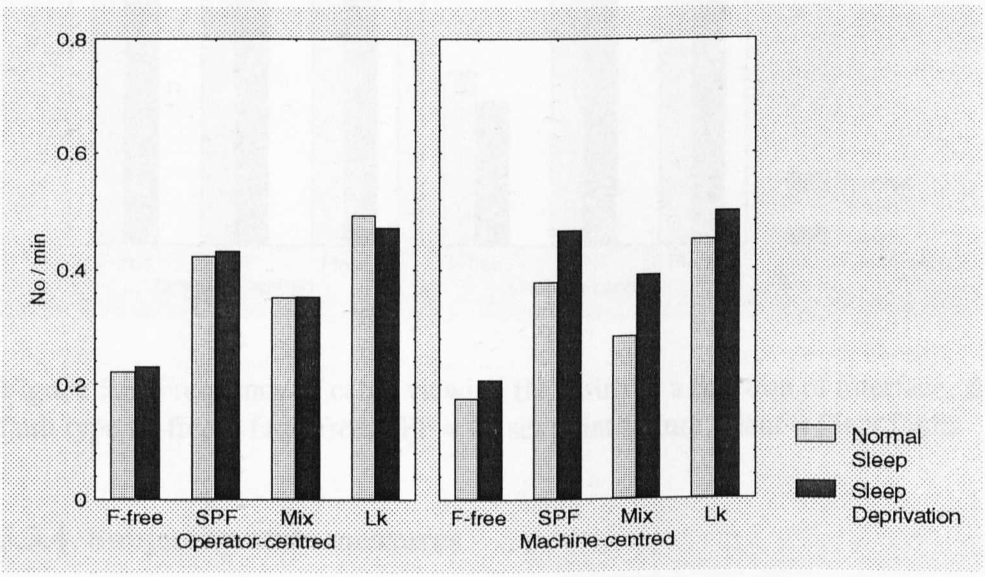


Figure 5.4: Frequency of control actions (No/min) as a function of interface, sleep and fault type (F-free = fault-free, SPF = O_2 set point failure, Mix = Mixer blockage, Lk = Cabin Leak)

System monitoring

The results showed that the cabin parameters were viewed more frequently under O-C than M-C (3.4 vs 2.5/min), which was statistically significant (F=14.64; df=1,15; p<.005). This difference was more pronounced under the fault-free phase than when a fault was present (see figure 5.5) as a result of the restricted access to information sources under M-C. This interaction between control and fault type was significant (F=9.45; df=2,30; p<.001). A main effect of sleep deprivation was found with more frequent system monitoring for non-sleep deprived participants (F=4.79; df=1,15; p<.05). System monitoring was reduced during fault-free periods (2.3/min) compared to

O_2spf (3.1/min) and mixer block/leak (3.3/min). The difference between fault-free and fault states was confirmed to be significant (F=67.7; df=2,30; p<.001; Post-hoc LSD-test: p<.001). Figure 5.5 presents the information sampling behaviour as a function of the main independent variables with the means for mixer block and leak again averaged under *plant fault* to stabilise the data.

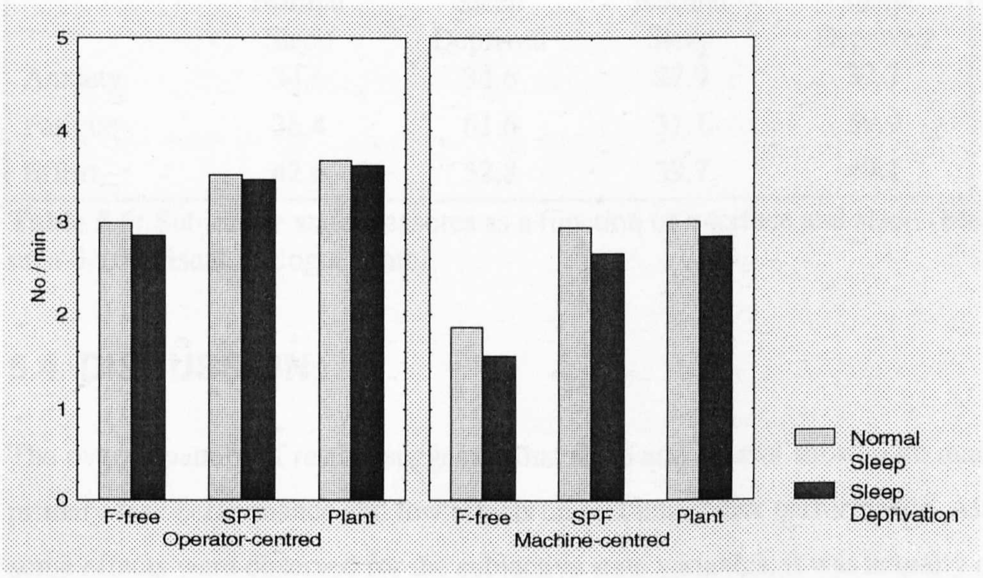


Figure 5.5: Frequency of cabin viewing (No/min) as a function of interface, sleep and fault type (F-free = fault-free, SPF = O_2 set point failure, Plant = Plant fault)

5.3.4 Subjective state measures

Anxiety

No significant main effects or interactions were observed for anxiety. Table 5.6 shows that anxiety levels were slightly reduced under M-C but the difference was not significant (F=2.11; df=1,15; p>.05). There was no main effect of sleep (F=2.39; df=1,15; p>.05) and no interaction (F=1.37; df=1,15; p>.05).

Fatigue

As shown in Table 5.6, fatigue was markedly increased as a result of sleep deprivation (F=70.59; df=1,15; p<.001). There was no significant effect of interface (F=2.68; df=1,15; p>.05) and no interaction (F<1).

Mental effort

A marked increase in subjective effort expenditure was observed (see Table 5.6) under sleep deprivation (F=12.30, df=1,15; p<.005). Again, there was no effect of interface (F=1.32; df=1,15; p>.05) and no interaction (F<1).

	Operator-centred		Machine-centred	
	Normal Sleep	Sleep Deprived	Normal Sleep	Sleep Deprived
Anxiety	34.6	34.6	27.9	32.7
Fatigue	36.4	61.6	31.1	59.9
Effort	42.6	52.8	39.7	49.1

Table 5.6: Subjective state measures as a function of interface and sleep (Mean ratings on a 0-100 visual analogue scale)

5.4 DISCUSSION

The overall pattern of results suggested that sleep and control showed no main effects on primary task performance and few effects on secondary task performance. Although some effects were observed for the subjective state variables, it was actually control actions and system monitoring behaviour that were most affected by the independent variables. Of particular interest was the interaction *sleep x interface*, where it was hypothesised that enhanced control would offset in part the negative effects of sleep deprivation. Evidence for this was found on some secondary tasks. The effects on the different dependent variable groups are now discussed in detail.

The indications were that sleep deprivation did not have any main effect on performance, neither for primary tasks nor for secondary tasks. However, system interventions were affected by loss of sleep, showing an increase in the overall number of control activities but at the same time monitoring activities were reduced. This suggested that sleep deprivation resulted in a stronger propensity towards action at the expense of system monitoring. Reduced system monitoring may be evidence of a riskier strategy, which consisted of relying more on alarms to be alerted to disturbances of the system state rather than checking information sources proactively. The reason for higher control activities may be that control strategies became simply less effective under sleep deprivation, resulting in more control actions being required to maintain performance. The lower level of system monitoring may have reduced the accuracy of the mental model of the current system state so that more control actions may have been necessary to stabilise the system. It is also conceivable that, because of the reduced monitoring activity in the sleep deprived group, the deviation of system parameters from the target states, when detected, had already been more advanced than under normal sleep so that more control actions became necessary. The increase in control actions under sleep deprivation was entirely accounted for by a higher use of the less resource-demanding *on/off* control actions, which were more indicative of feedback (or closed-loop) control (e.g. Wickens 1992). *Setting* control actions, on the other hand, did not increase. They correspond more to feedforward (or open-loop) control, which requires the operator to make more precise predictions about the consequences of control actions to arrive at the target state. Effective feedforward control requires a better general knowledge of the system and its current state, and more cognitive-energetical resources during the control process than closed-loop control. As a result, in real process control operators were often found to shift from open-loop control to closed-loop control when task demands began to increase (Bainbridge 1974).

Despite the unfavourable operator state caused by sleep deprivation, no direct performance decrements were observed as a result of increased fatigue levels. Instead, the results suggested that, in addition to adaptive changes in information sampling behaviour, increased effort expenditure was used as the primary strategy to combat the effects of sleep loss.

It was hypothesised that reduced dialogue control would have negative effects on performance because it restricted the use of preventive control actions, and the lack of system feedback made the maintenance of an up-to-date mental model of the system state more difficult. Nevertheless, most performance measures did not show any evidence of decrements under M-C. Only for alarm acknowledgement was an increase in response time observed as a result of the restricted interface. This effect was even more pronounced under sleep deprivation. This may suggest that less involvement in the control process under M-C led to longer RT since more time was needed to verify the alarm. However, there was also some evidence from post-experimental interviews that increased RT under M-C was the result of a deliberate strategy of some operators to regain control over the system. Participants admitted having disobeyed experimental instructions by not responding to the alarm immediately because the presence of an alarm state allowed them to interrogate and manipulate all system components without restriction. This corresponds to a finding previously found in research in automated systems, where operators deliberately violated rules in order to (re-)gain control over the system (Hockey and Maule 1995).

The absence of primary task performance impairment under M-C was unexpected (the data even showed slightly better performance under M-C) because of the acknowledged advantage of a user-centred design. However, a closer analysis of the session logs revealed that the permanent access to the system under O-C had tempted many participants to explore the system by taking over manual control when it was not required. This led in some instances to parameters moving out of their target state, solely as a result of system exploration and not brought about by any system failure. Reducing unnecessary and counterproductive operator intervention is exactly the reason why engineers have designed automatic systems with limited operator access. Should we therefore conclude that this design philosophy is fully justified? System exploration is generally to be considered as a positive behavioural pattern since it helps to refine the mental model (see also Chapter 6) and is therefore likely to have long-term effects for skill improvement. However, it may bear the risk of disturbed system states being caused by sub-optimal manual control behaviour but this risk appears to be acceptable considering the potential benefits of an interface with enhanced dialogue control. The use of unnecessary manual interventions may also have been affected by the lack of trust in the system (Muir 1988) or over-confidence in one's control skills (Kleinmuntz 1990, Moray and Lee 1990). Although none of these hypotheses can be formally tested in this experiment, they are likely to have influenced the scheduling of manual control activities. Experiment 2 will explore in more detail the issue of system management strategies by attempting to influence knowledge structures and system management behaviour by means of different training methods. In this context it would also be interesting to determine whether system exploration occurs primarily during the familiarisation process with the system (what we would expect) and becomes less pronounced (or even disappears) with increasing experience of the operator. The field experiments reported in Chapter 7 (in particular Experiment 6) may provide us with an answer since participants completed a large number of testing sessions (up to 60 hours).

The combination of sleep loss and an M-C interface was thought to be particularly vulnerable since in addition to reduced efforts to sample information sources, the interrogability of the system is restricted, which makes it even more difficult to maintain an up-to-date mental model of the current system state. The interaction became evident in two measures. The best diagnostic performance in the fault identification task was recorded under the most favourable condition (Sleep and O-C). This may suggest that unrestricted access to the interface is a prerequisite for effective diagnosis since it provides the operator with the possibility of engaging in a diagnostic process at any time. This permits the operator to determine system failures well before they have caused disturbances. However, only the alert operator benefited from enhanced control. This might be because sleep deprived operators did not have sufficient resources available for effective fault diagnosis so that they did not better than operators in the M-C conditions. The second measure was reaction time, where the delayed alarm acknowledgement under M-C (compared to O-C) was much more pronounced for the sleep deprived operator.

As predicted fault difficulty had an effect on primary performance measures but no interactions with any of the other IVs were found. The absence of interaction effects was unexpected since it was hypothesised that effects of sleep deprivation and restricted dialogue control were more likely to manifest themselves under increased fault difficulty. However, secondary tasks decrements (such as prospective memory errors and RT) did occur with increasing fault difficulty and they were indicative of the higher resource demands required for the primary task under these conditions. This was in line with the theoretical framework of the thesis, which predicted secondary task performance to suffer more strongly than primary task performance

To interpret the results in the context of the methodological framework, it appeared that increased effort expenditure and adaptive changes in control and information sampling strategies were the main adaptive responses observed as a function of sleep loss. For dialogue control, no evidence of *compensatory costs* as an adjustment pattern were observed, but information sampling behaviour and a small number of secondary tasks were affected. The small number of secondary task decrements indicated that participants attempted to maintain performance on all tasks rather than focusing selectively on high-priority tasks.

Chapter 6:

Knowledge and Training

6.0 Summary

This chapter reports two lab-based experiments that examined the effect of training on short-term and long-term skill retention. The research literature suggests that the effectiveness of a particular training method is difficult to assess without considering the particular circumstances the knowledge is applied to. It is even more difficult to answer the question of what training method is most resistant to skill decrement after an extended lay-off period. Hardly any research has examined this question despite its vital importance to extended spaceflight.

The first experiment examined the effect of two training approaches (procedure-based vs system-based instructions). After extensive training on the task (6 hours), 25 participants were tested during an extended 3-hour experimental session, which included noise (as a prominent space stressor) and fault difficulty as within-subject variables. The results showed few effects of training on performance. However, an analysis of the mental model structure, and of control and information sampling behaviour revealed considerable differences between training groups. Noise had an effect on information sampling strategies and on the most vulnerable secondary task (prospective memory), whereas performance on primary tasks remained unimpaired. The findings suggest that the different training approaches brought about differences in the underlying knowledge structure. While these may not manifest themselves in performance under normal conditions, they are more likely to emerge under more extreme environmental and operational circumstances.

The second experiment, carried out 8 months after the first, examined the effects of an extended lay-off period on skill maintenance. The same participants (N=17) were tested a second time without having had any practice on the task during the intervening period. Despite the long retention interval, there was little overall evidence of skill decrement. On the contrary, some performance measures even showed improvements. The results also indicated that the differences observed between training groups at the earlier testing time were considerably less pronounced on the later occasion. The finding suggests, first, that skill decrement as a result of an extended lay-off period might be of less concern (at least for the cognitive tasks examined) than previously thought and, second, that any advantage gained from using a certain training method may be at risk of vanishing during an extended lay-off period.

6.1 Introduction

6.1.1 Training and performance

The fact that training cosmonauts can take up to 10 years (Bluth and Helppie 1987) indicates the enormous amount of preparation necessary to obtain an adequate level of skill for the successful completion of a flight. In the Soviet space programme, training was divided into four main areas (Bluth and Helppie 1987): (1) Preparation for spacecraft control, operation and maintenance of on-board systems, (2) Performance of scientific experiments, (3) Preparation for the effect of spaceflight factors, and (4) Psychological training. The issues addressed in this thesis refer to the first area, which covers the operation and maintenance of the life support system. Since this is essentially an automated process control system (see section 2.2), it is appropriate to focus the review on the training literature on process control. It is centred on the question of what type of training is needed to achieve good performance in all aspects of system control and maintenance.

Four different training approaches have traditionally been used to teach operators trouble-shooting strategies for the identification and rectification of system faults (Morris and Rouse 1985): (1) Instruction in the theory upon which the system is based, (2) Provision of opportunities for trouble-shooting practice, (3) Guidance in the use of system knowledge, and (4) Guidance in the use of algorithms or rules. There is some evidence to suggest that operators who had been given procedural instructions did not show poorer performance (sometimes they did even better) than operators whose training was based on the teaching of system principles (e.g., Shepherd, Marshall, Turner and Duncan 1977; Crossman and Cooke 1974). This did not only apply to familiar faults (i.e. those which were discussed or practised during training) but also to novel faults, that is, those encountered for the first time in the experimental session. However, another experiment showed that subjects trained on diagnostic heuristics performed less well on novel faults (but did equally well on familiar faults) than a second group, which received a 'technical story' about the functioning of the system components (Patrick and

Haines 1988). Performance improved significantly if it was made explicit how to use the theoretical knowledge in applied problems connected to the system (Morris and Rouse 1985). In several studies, it was found to be beneficial for performance if procedures were provided to guide the trouble-shooting approach of subjects (Duncan 1971; Goldbeck, Bernstein, Hillix and Marx 1957). In these tasks a faulty element in a network of interconnected components had to be identified by keeping the number of tests carried out in the process to a minimum.

Repeated practice has been identified as one of the most important factors to enhance trouble-shooting performance (Morris and Rouse 1985). Furthermore, action-related feedback is thought to play a substantial role in improving trouble-shooting performance (Patrick 1992).

The review of the literature identified two problems concerning the relevance of the material for this thesis. First, although some of the studies were conducted with the goal of providing new insights into complex problem-solving in process control situations, the vast majority of them employed non-dynamic tasks (i.e. the operator needed to find a fault in a static situation). Second, a considerable number of studies compared an experimental group (receiving specific instructions) with a control group (no specific instructions) instead of using ecologically more relevant approaches (e.g. procedural vs knowledge-based instructions), as was the case in the study by Patrick and Haines (1988).

6.1.2 Knowledge representation

The representation of knowledge is closely linked to the two issues discussed in the preceding section: Performance and Training. It can be regarded as an intermediate factor in the causal link between training and performance with training influencing knowledge representation, which in turn influences performance. These knowledge representations have been referred to as *mental models* in a number of different contexts (Craik 1943; Johnson-Laird 1993; Norman 1983). While Johnson-Laird investigated the issue in the context of deductive reasoning, this thesis is more concerned with the other strand of research, which looks at the representations of physical systems in the human mind. The primary goal of that work was to obtain a better understanding of the behaviour of humans when operating different kinds of physical systems. Inaccurate mental models of the system, or its current state, have been blamed for a number of serious incidents (e.g. accident at Three Mile Island; see Reason 1990), thereby more than justifying continued research efforts in that area.

Mental models have been described along a number of different dimensions, for example, knowledge level and knowledge categories (Wirstad 1988). The first dimension refers to the level of accuracy and elaboration of the mental model. With the process of mental model refinement being regarded as a continuous activity, the mental model changes as a result of training or through the interaction of the individual with the system (De Kleer and Brown 1983). During the knowledge acquisition process, the individual makes a number of implicit assumptions about the functioning of the system, which are constantly modified if inconsistencies between the model and the external reality emerge. This adaptation process leads to continuous improvements in accuracy and an increase in complexity of the mental model. However, some authors pointed out that in practice the process of continuous refinement may be hampered by a number of factors. Individuals were observed to ignore information that challenged their (inaccurate) model of the situation by actively seeking information which confirmed their already chosen hypotheses (Mynatt, Doherty and Tweney 1977). Even if the evidence for inconsistencies in a model becomes overwhelming, operators may still be reluctant to change their model. This phenomenon has been called *confirmation bias.* Furthermore, mental models tend to be rather unstable, with individuals forgetting details about the system if they have not used it for some time (Norman 1983). Individuals also have a strong tendency to simplify the model with the aim of reducing cognitive complexity. As a result of the interaction with the system, the individual may also develop superstitious behavioural patterns to manage the system (Norman 1983). This undesired effect is, however, more likely to occur during unsupervised practice than under the guidance of a trainer. To overcome the problems associated with the development of an inaccurate mental model, appropriate training is generally considered as the most promising means to change the structure and content of a mental model (Wickens 1992).

The second dimension (knowledge categories) refers to the different aspects of system knowledge. A classification system has been proposed to distinguish between different types of knowledge of the physical and functional aspects of the system (Wirstad 1988).

It essentially covers the following aspects of system knowledge: *plant layout and components*, *cause-effect relationships between process parameters*, *constraints of operation* (e.g. max. allowable levels of parameters) and the *current state of the process* (may also include state in recent past and in near future). This asserts the notion that mental models are multi-faceted, with some facets referring to constantly changing knowledge (such as current state of the process) whereas others are of a more stable nature. This suggests that the stable knowledge structures are primarily determined by the type of training received. The quality of knowledge of the current system state, on the other hand, is more likely to be a function of system design features such as transparency, feedback and interrogability.

To summarise the evidence from the literature, mental models of physical systems are considered an important determinant of performance in the management of these systems. As a result, training has become an issue of primary concern since it is considered an important means to manipulate the mental model structure.

6.1.3 Skill retention

The unique characteristics of extended spaceflight as a work environment raise a further problem regarding the retention of skill for low frequency tasks. Extended spaceflight comprises distinct mission phases (e.g., launch, cruising, docking, landing), with each of them requiring specific skills that may not be needed during any other phase. This means that skills associated with operations during a short and infrequent event (e.g. docking) may not be practised for long periods of time. For those tasks, the risk of skill decay is increased. One may argue that the availability of computer-based simulations, embedded in the operational environment, may provide sufficient possibilities to maintain any skills under threat. However, in order to design an optimal on-board training programme, one must first identify the skills most at risk in order to allocate the necessary training time for their maintenance. Otherwise, one's efforts may result in a training programme that does not reflect the level of skill decrement associated with each task. Furthermore, once the mission is under way, any identified skill degradations are more difficult to deal with than in comparative terrestrial work environments. In the latter situation more options would be available to tackle the problems (e.g., exchanging operators, organising a refresher course). Therefore, spaceflight requires very conscientious planning of in-mission training since the possibilities for change are limited during the mission itself. The problem of maintaining skill levels on low frequency tasks is most serious for those tasks that are also 'non-forgiving'. This means the crew member needs to get it right at the first attempt since a repetition of the procedure is not possible (e.g., during certain docking manoeuvres, planetary landing).

A review of the literature did not reveal a large amount of research in the field of skill retention. This was confirmed by an earlier review carried out by Hagman and Rose (1983) who concluded that there is a shortage of psychological research into the issue of skill retention. Although there is a large vein of research in the field of learning (e.g. Houston 1986), it is only of limited use in the present case. This is because short time intervals were commonly used to measure effectiveness of training, which does not address the problems faced in extended spaceflight. However, some issues associated with long-term skill retention were addressed in a military context (Rose 1989). This was because the military observed problems with the retention of crucial skills after extended lay-off periods.

Most studies that made use of longer retention intervals focussed on perceptual-motor skills and the retention of verbal material. The general picture suggested that perceptual-motor skills on continuous tasks (e.g. tracking) were well retained even after long time periods (Annett 1989, Patrick 1992). One study used a retention interval of 24 months for a tracking task and found very little loss of skill indeed (Fleishman and Parker 1962). However, the picture looked very differently when discrete tasks (involving step-by-step procedures) were examined. They were clearly less well remembered, with retention deteriorating as a function of the number of steps involved (Hagman and Rose 1983). The retention of verbal material also appeared to decline quite rapidly (Annett 1989). The results of a study of Goldberg and O'Rourke (1989) indicated that type of training did not show any long-term effects of skill retention after performance of both groups had reached equal skill levels during training. This means that certain forms of training may initially lead to superior performance. However, once the alternative training method has achieved the same performance level (albeit with more training time), no long-term benefits of the superior training method are to be expected.

From our point of view, the weak points of the studies reported in the literature is that most employed tasks with little ecological validity, following the old tradition of single task experiments (see also Johnson 1981). Furthermore, most studies used rather short-term intervals to assess skill decay, which did not reflect the time scale of extended spaceflight. One obvious reason for the paucity of research in this area is that longitudinal studies of this kind are difficult to conduct since they require the maintenance of a participant pool for long periods of time. An earlier review of the literature specifically addressed the problem of skill retention in spaceflight (Gardlin and Sitterley 1972), but it concluded that the tasks used in previous research were totally unrepresentative of space tasks.

6.1.4 Noise

Although space is not noisy in itself, the space traveller is confronted with a permanent level of noise caused by the machinery of the spacecraft. There are two types of noise-related problems during spaceflight (Connors et al 1985). The first one refers to short-duration peak levels (between 120 and 130 dB) occurring during launch and re-entry phases. The second problem concerns the continuous noise during the cruise phase, which is an issue particularly relevant to extended spaceflight. Over the years, a considerable reduction of noise levels has been achieved from levels of around 70 dB (Apollo flights) to approximately 43 dB in the living quarters of Skylab (Connors et al 1985). However, even these reduced levels may still be too high for extended spaceflight. The requirements for Soviet flights were even stricter, aiming at maximum levels of 40 dB (30 dB during the night) for living quarters and 65 dB for working quarters (Bluth and Helppie 1987).

Four aspects may be considered when looking at the issue of noise. First, high levels of noise may cause damage to the ear, resulting in a temporary or even permanent threshold shift (Jones 1983). However, this could only occur under very high levels of noise (e.g., during launch and re-entry phase) and is therefore not a prominent feature of extended spaceflight. Second, it may represent a physiological stressor, resulting in adverse effects on various bodily systems such as the cardiovascular, the vestibular and the autonomic nervous system (Connors et al 1985). Third, it may affect various aspects of performance. Fourth, noise is generally regarded as causing fatigue due to a more rapid depletion of cognitive resources during noise exposure (Jones 1983).

This review of the noise literature focuses on the latter two aspects (performance and fatigue) since they are more relevant to the psychological features of extended spaceflight. Continuous noise is the predominant type of noise during a space mission, caused by the life support systems that are in operation 24 hours a day.

The findings in the literature on continuous noise are not entirely consistent concerning its impact on performance for different cognitive tasks. A number of studies looking at monitoring tasks generally indicated an impairment under noise. Although noise did not show any effect in a signal detection task using a single source (e.g. Blackwell and Belt 1971), it affected performance (as demonstrated in a series of experiments by Jerison 1957, 1959) when the number of information sources to be monitored was increased. There was also some evidence suggesting that when confronted with several information sources, humans tended to sample more frequently from sources with a greater probability of signal occurrence, independently from the spatial location of the source (Hockey 1970a, 1970b). Broadbent (1976) reported that under noise exposure humans showed higher confidence in stating whether a signal was present or not, basically displaying more risky decision-making behaviour. In serial reaction time tasks, individuals were found to operate with reduced accuracy during noise exposure while response speed was maintained (e.g., Wilkinson 1963), or in some instances even increased (Davies and Davies 1975). If verbal information is to be processed, noise already begins to impair performance at lower sound levels than commonly observed for other cognitive tasks (Weinstein 1977). It is generally assumed that working memory is a key factor, which determines the extent to which performance suffers as a function of noise (Hockey 1986). This is because noise reduces the information processing capacity, resulting in performance impairment in tasks sensitive to working memory. It appears that working memory is a critical factor in determining the impact of noise. Another study (Dörner and Pfeifer 1993), using a complex and dynamic decision-making task (see Chapter 4 for a task description), showed that noise did not impair overall performance levels (there was even a small but non-significant performance improvement under noise). Instead, it was found to affect the strategies employed to manage information. Dörner and Pfeifer concluded that stressed subjects (those exposed to noise) worked with a lower level of resolution, which facilitated the detection of the right priorities instead of being distracted by insignificant information. The factor 'time-on-task' appears to play an important part in the prediction of noise-induced performance decrements. This is particularly relevant in the context of extended spaceflight since noise exposure is virtually permanent, though at a reduced sound level. The evidence suggests that any negative impact of noise generally increases with time-on-task (Wilkinson 1963). On the other hand, noise was also found to help maintain performance levels on some tasks such as pursuit tracking (Hockey 1970b). It may be that the more demanding the task in terms of higher cognitive resources, the more likely it is to suffer as a function of exposure time.

Since it is impossible to withdraw from the noise source during a space mission, the control aspect is likely to play an important role. Lack of control over a noise source tended to result in higher stress levels than situations with the possibility of control (Cohen, Glass and Phillips 1979). This applied even in situations, where the possibility of reducing noise levels was never used. On the other hand, if noise is associated with one's safety or well-being (which would surely apply to life support systems), stress levels are reduced (Sader 1966). Bluth and Helppie (1987) reported anecdotal evidence from the Soviet space programme with cosmonauts subconsciously listening to machinery during the night to monitor the correct functioning of the systems.

With the issue of extended spaceflight in mind, the review of the noise literature raises two problems. First, most of the studies have used noise levels of between 70-115 dB, which would exceed any level space travellers would be exposed to during the cruise phase. Not many experiments have been conducted, which used long-term exposure to moderate noise levels. Bluth and Helppie (1987) reported a seven-month isolation study, in which three Soviet scientists were continuously exposed to noise. The results indicated that there was no adaptation to noise in that it was regarded as very unpleasant throughout, in particular during the rest periods. Some other Soviet studies demonstrated the cumulative effects of noise during prolonged exposure (Bluth and Helppie 1987). Some correlation studies looked at the effects of prolonged exposure to noisy environments (e.g. factories, airports) and indicated the occurrence of physiological and psychological changes, such as hypertension and irritability (Jonsson and Hansson 1977; Cohen, Evans, Krantz and Stokols 1980). Despite the availability of a small number of studies of long-term noise exposure, one mainly has to rely on extrapolations from studies using short-term exposure to high noise levels. The findings are supplemented by anecdotal evidence from real spaceflight, where noise was identified as an impediment to effective communication (Kelly and Kanas 1992).

The second problem is that cognitive tasks used in the mostly lab-based experiments showed considerable differences to real crew tasks, mainly with regard to the degree of complexity. Only more recent studies (e.g., Dörner and Pfeifer 1993) employed task environments of higher complexity with multiple tasks to be accomplished.

One may conclude that the impact of noise on a multiple task battery is difficult to predict. This is because humans tend to deploy their resources strategically and focus on high-priority tasks while neglecting less important tasks, with the goal of overall performance maintenance (Jones 1983; Hockey 1993). Despite the performance-enhancing effects of noise found in certain contexts, noise still needs to be regarded as a stressor that may pose a threat to performance maintenance on long missions.

6.2 Experiment 2

The primary goal of this experiment was to evaluate two training approaches in combination with noise and different fault types. Concluding from the literature review on training in process control, it seems not to be a trivial question to ask what kind of training would be most successful in a particular context, because there are certain trade-offs associated with each approach. It was predicted that procedure-based training would be superior if operators had to deal with familiar fault states, since a rule-based approach was considered to be more effective. The system-based training approach, on the other hand, was expected to show better results in unfamiliar fault situations when real-time problem-solving was required. Furthermore, it was hypothesised that noise, as a stressor reducing working memory capacities, would affect a rule-based system management approach to a lesser extent than the knowledge-based strategy. In addition, non-priority tasks (in particular those relying heavily on working memory) were predicted to be most vulnerable to decrements as a result of noise exposure. Changes in system management strategies were also expected to occur during the presence of noise, and as a result of increasing fault difficulty, but the direction of the effect was difficult to predict.

6.2.1 Method

6.2.1.1 Design

The experiment employed a repeated measures design (ABA) with a between-subject variable (training) and two within-subject variables (noise and fault type). Training was varied at two levels (system-based vs procedure-based), noise at 3 levels (quiet, noise, quiet) and fault type at 4 levels (faultfree, practised faults, novel faults, control panel failures). Details of these are given below. Using AMT 3.0 as the experimental task, the experimental session lasted for 3 hours, with each 1-hour period associated with a different noise condition. In each 1-hour period, participants encountered fault types with the following frequencies: three practised faults, two novel faults and one control panel failure. Following an experimental plan, the order and duration of faults was balanced across fault types. Fault phases lasted between 300 and 660 secs, with fault-free phases occurring at the beginning of each period and between fault states. Participants did not know the order in which faults would occur nor their duration. The following example illustrates a typical set-up of a 30-min sub-period: Fault-free (min 0-3.5), Practised fault (min 3.5-11), Fault-free (min 13-15), Control panel failure (min 15-22), Fault-free (min 22-24.5), Novel fault (min 24.5-29.5).

6.2.1.2 Participants

Participants were mainly recruited among the Hull University student population. They were chosen on the basis of some basic selection principles to ensure that they matched the elementary criteria for the selection of trainee astronauts. They were required to be computer literate as well as to have some basic knowledge of the functioning of process control systems. Furthermore, the participants should be pursuing or have completed their studies in either a pure or applied science. They were all invited to an informal interview to ensure that they fulfilled the selection criteria and to give them some more detailed information about the experiment.

As a result of the selection procedure, 25 participants (23 male, 2 female) were recruited, all members of the University of Hull either as undergraduates (8), postgraduates (14) or staff (3). The participants were members in (or graduated from) the following departments: Engineering (15), chemistry (7), biology (2) and computer science (1). Their ages ranged from 20 to 30, with a mean of 24.2 years. Because of the high commitment of the participants, all 25 successfully completed the training phase and the subsequent experiment.

6.2.1.3 Training

General structure of training

With type of training being one of the independent variables, participants were randomly assigned to two different training regimes: *procedure-based* and *system-based.*

Thirteen participants received *procedure-based training*, which meant that they were extensively instructed about precise procedures to follow when a particular type of fault occurred. Twelve participants were given *system-based training*, which encouraged them to understand the relationship and interactions between the different components of the system. During the occurrence of faults, they should base interventions on their understanding of the causes of the disturbance rather than following a guide of step-by-step procedures.

Both groups received approximately 6 hours of training, which comprised 5 sessions. The first session was identical for both groups, in which the basic functioning of AMT was explained. In the following four sessions, both groups were trained on exactly the same system faults but with a different emphasis (see below). The instructor occasionally interrupted the session (via a pause button) to give explanations while the system state was temporarily frozen. This had the advantage that after the explanation had been given the session could be resumed (via a restart button) without the participant being faced with a completely changed situation. The training approach corresponded to a *closed-loop training system with continuous progress monitoring* (Patrick 1992).

Both groups had a fault finding guide at their disposal, which they could refer to during training as well as during the experimental session. The guide was problem-centred and provided quick access to the identification of faults and the appropriate control strategies to deal with them. A copy of the fault finding guide can be found in Appendix 2. The contents of each of the five training sessions are now outlined.

Training sessions

(1) In the first session the system was introduced, explaining the possibilities of controlling and interrogating the different functions. Furthermore, participants were instructed in primary and secondary tasks, which they had to complete during system operation. At the end of the session they were allowed to try out the controls of the system and asked to complete some of the tasks.

(2) The second session began with a brief recapitulation of the main features of the system. The fault finding guide was then introduced, explaining the symptoms of all system faults and their most effective control strategy. Participants went through two twenty-minute practice trials, during each of which the same three faults occurred.

(3) In the third session, the participants completed three twenty-minute practice trials, in which three new faults were progressively introduced. The trials were interrupted from time to time (via the pause button) to give participants the opportunity to consult the fault finding guide to check the symptoms occurring on the screen.

(4) At the beginning of session four, participants were given time to study again the fault finding guide. They then completed another two twenty-minute trials comprising the faults they had encountered previously. They were also told that other faults might arise during the experimental session but they would not be able to practise them.

(5) The structure of the last session was similar to the previous one. However, some of the trials were completed under noise exposure to familiarise the participants with the headphones to be used in the experiment, and to ensure that the noise level was acceptable to them.

Procedure-based training

The participants were trained on each fault by emphasising the importance of following the procedures for fault identification and management, both of which were described in the fault finding guide. As a result, participants in this group generally spent more time consulting the guide than the system group. After participants became more familiar with the different faults, they operated the system more and more independently. However, if they did not follow the prescribed procedures, the instructor intervened and again outlined the steps to be taken. During procedure-based training, no explanations were given about interactions between the different subsystems or how a particular fault may affect the principal system parameters. If a participant asked a question concerning these issues, the importance of following the procedures was stressed as being the most efficient way of managing the system.

System-based training

The training sessions for the system group were similar with regard to the sequence of faults practised. However, one difference was that the faults were grouped according to types (e.g., leaks, valves stuck open) and introduced at the same time. This was thought to facilitate the understanding of the effects of each fault on the main system parameters. This was in line with the general idea behind the training regime, which was to deepen the participants' understanding of the closely-coupled subsystems and how they affect each other. In practice this was done by asking the participants about their expectations of the symptoms of a particular fault (i.e. effects on different parameters) and why they held these particular views. They were constantly encouraged to make a priori predictions about system behaviour and to give a posteriori explanations. The instructor challenged their views if they were inconsistent or did not reflect the actual processes in the system. The ultimate goal was to refine their mental model of the system. It was predicted that this would enable them to cope more easily with previously unencountered system faults.

6.2.1.4 Noise

As an important feature of the ambient space work environment, noise was one of the main independent variables. It was varied at two levels, with a repetition of the noise-free condition in the last period (ABA design). This meant that the first 1-hour period was free of noise. During the second 1-hour period noise was applied through a set of headphones, followed by another 1-hour period without noise. The noise reached levels of approximately 80 db(A) and was generated by different types of machinery (e.g. drills, power saws). This is higher than current noise levels in space vessels (only launch and re-entry produce higher levels) because it was attempted to balance out the short exposure time (only 1 hour) by increasing the intensity of the stimulus. This corresponds to a commonly accepted law in noise research, postulating that stress effects are a function of exposure time and stimulus intensity.

6.2.1.5 Fault difficulty

Four levels of the fault variable were manipulated in the experiment. (1) In the *faultfree (F-free)* state the automatic controller operated perfectly well and no system component failed or was defective. (2) *Practised faults (PracF)* were fault states that were extensively practised in training. (3) *Novel faults (NovF)* referred to faults that had not been encountered before but were of the same general type as the practised faults. (4) *Control panel failures (CtrlPanF)* were also unpractised faults but of a different kind to those encountered in training. The operator had to develop an alternative strategy since the most straightforward approach of dealing with the fault (via the corresponding control panel) was unavailable due to an additional failure of the needed control panel. Furthermore, the maintenance facility was not in operation for the duration of the fault, hence the operator was unable to repair the fault. It was expected that control panel failures would be most difficult to manage, followed by novel faults and practised faults.

6.2.1.6 Dependent Variables

As described in Chapter 4, the AMT software permits the investigation of a large number of variables since it records any action taken by the operator during system operation. In line with the study rationale, the following domains were covered in the analysis: *Primary task performance, secondary task performance, system interrogation and intervention, subjective operator state* and *mental model of operator*. The dependent variables from each domain are described in more detail below.

Primary task performance. (1) *Parameter control failures* referred to the percentage of time each of the five key parameters had been out of range (i.e. in the red zone). (2) *Diagnostic speed* was scored as the time needed to carry out a correct fault diagnosis. (3) *Diagnostic accuracy* provided an index of the number of incorrect attempts to diagnose a fault.

Secondary task performance. (1) *Annunciator acknowledgement* was a reaction time measure to indicate the time needed to respond to an alarm signal. (2) *Tank level recording* was a prospective memory task, where recordings were scored for accuracy of timing. (3) *Resource conservation* referred to the amount of oxygen and nitrogen consumed during system operation.

System interrogations and interventions. (1) *Control actions* referred to the frequency of system interventions. (2) *History display sampling* measured the frequency of interrogating the historical display. (3) *Control panel sampling* measured the frequency of sampling the control panel without intervention. (4) *Flow meter sampling* was an index of the frequency of viewing the flow meters. (5) *Interrogation of maintenance facility* provided a measure of the frequency of accessing the maintenance facility without carrying out a repair.

Subjective state. The following three subjective state measures were collected after each working session: (1) *Anxiety,* (2) *Fatigue,* and (3) *Mental effort.*

Mental model. (1) Explicit knowledge referred to the system knowledge demonstrated in a post-experimental questionnaire. (2) *Implicit knowledge* referred to the system knowledge demonstrated during system control activities.

6.2.1.7 Procedure

The following four practised faults were used during the experimental session: *Leak of oxygen valve, Leak of mixer valve, Vent permanently open* and *Nitrogen valve permanently open.* The four novel faults presented in the session were: *Block of oxygen valve, O_2 set point failure, N_2 set point failure* and *Dehumidifier set point failure.* The following three control panel failures were used: *Cooler failure* (with a simultaneous temperature control panel failure), *N_2 valve permanently open* (with a simultaneous pressure CtrlPanF) and *O_2 valve permanently open* (with a simultaneous O_2 control panel failure). Two faults were practised during the training sessions but did not occur in the experiment: *Carbon dioxide scrubber failure* and *cooler failure.*

6.2.2 Results

This section reports descriptive statistics and the results of the statistical tests. For most performance data, a square-root or log-transformation was carried out to reduce the skewness of the distribution. This is a standard procedure applied to data, which are often positively skewed when performance is expressed in error terms. However, for reasons of clarity the means presented in figures and tables represent the original values.

6.2.2.1 Primary task performance

Parameter control failures

All key parameters showed a strong effect of fault type with the number of control failures increasing with fault difficulty (see Figure 6.1). The highest error score appeared during CtrlPanFs, followed by NovFs and PracFs. No control errors were committed during the F-free period. The effect was highly significant (F=514.2; df=3,69; p<.001). Post-hoc LSD tests showed that all four means were significantly different from each other (p<.05). The system group (PCF: 3.5 %) did slightly better than the procedure group (PCF: 4.1 %) although the difference was not significant (F=3.7; df=1,23; p>.05). Better performance of the system group was only observed for the more difficult fault states (NovF and CtrlPanF) but not for the practised faults. This corresponded to the predictions made, which assumed a superiority of the system-based approach for unfamiliar fault states. As hypothesised, the system group showed some deterioration of performance under noise, whereas the procedure group did not show this decline. However, the effect was not significant (F=1.5; df=2,46; p>.05). No significant interactions were found.

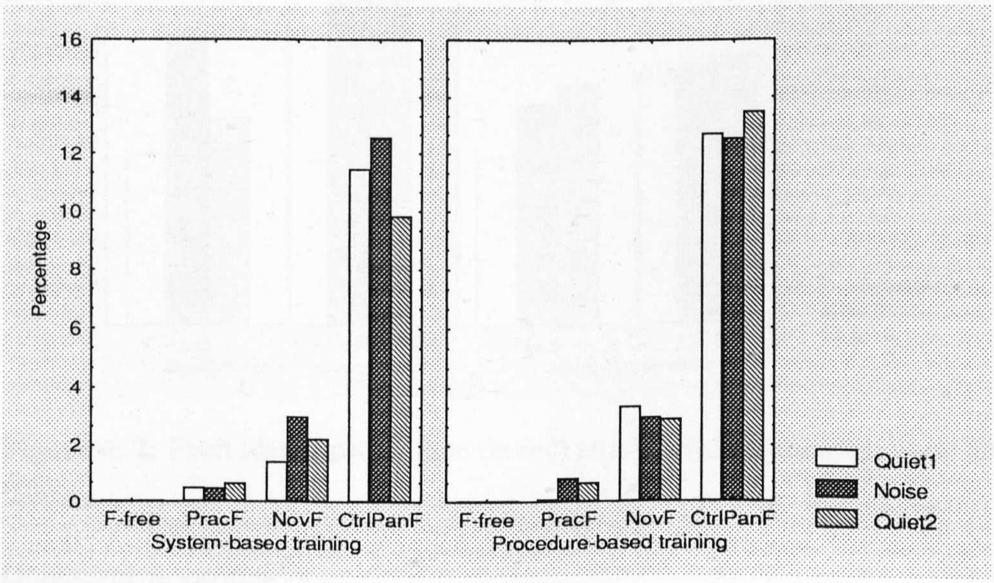


Figure 6. 1: Mean parameter control failures (%) as a function of training, noise and fault type

Diagnostic speed

The efficiency of fault diagnosis was evaluated by analysing the time needed for a correct diagnosis. Control panel failures were not included in this analysis. They could not be diagnosed because the failure of the maintenance facility was part of the fault state. The maximum value for fault identification time was set to 300 sec to control for differences in the duration of fault states. As expected, fault identification was quicker for PracFs than NovFs (95 sec vs 135 sec; F=23.1; df=1,23; p<.001). Procedure-based training was found to be superior to system-based training with regard to fault identification times (F=5.96; df=1,23; p<.05). The difference between the training groups was due to an unexpected interaction between fault type and training (F=4.73; df=1,23; p<.05), with the system group identifying novel faults considerably more slowly than the procedure group (158 sec vs 115 sec). No difference between training groups was observed for the practised faults. No significant effect of noise was found although identification time increased during noise exposure (F=2.94; df=2,46; p>.05). This increase was accounted for by the system group but again the interaction effect (training x noise) was not large enough to reach significance (F=1.25; df=2,46; p>.05). Fault identification times for all conditions are presented in Figure 6.2.

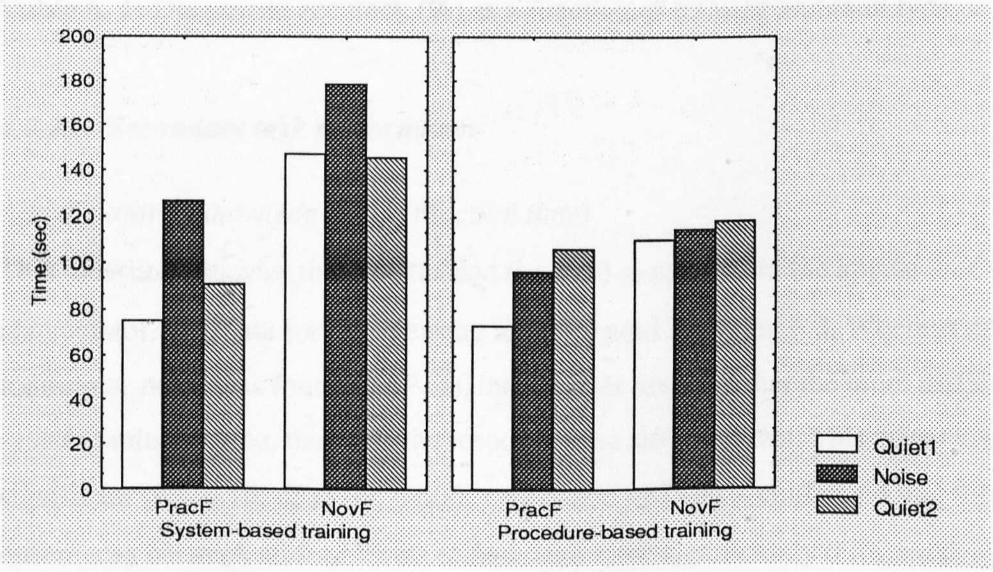


Figure 6. 2: Fault identification time (in sec) as a function of training, noise and fault type

Diagnostic accuracy

This measure was scored for the percentage of correctly diagnosed faults. In contrast to the diagnosis time, an analysis of the accuracy of the diagnostic process revealed no significant difference between training groups. Most faults were correctly diagnosed at the first attempt in both groups (overall mean: 81.8%). A further 11.3 % of faults were correctly diagnosed at a later attempt, leaving 6.9% of faults where a correct diagnosis was unsuccessful. However, there was a trend for the procedure group to perform slightly better than the system group (93.4 % vs 92.8 % correct diagnoses) but the statistical test failed to confirm the significance of the difference (F<1). A comparison between fault types revealed that PracFs (95.6%) were more accurately diagnosed than NovFs (90.6%). This difference was significant (F=4.64; df=1,23; p<.05). For both criteria, the effect of noise and interactions were not significant (all F<1). Using the stricter criterion (correct diagnosis at first attempt), a second analysis revealed the same pattern of results but none of the effects were significant. Table 6.1 shows the percentage of correct diagnoses (first or later attempt), averaged across periods, for training and fault type.

	System group	Procedure group	Total
PracF	95.4	95.8	95.6
NovF	90.2	91.0	90.6
Session	92.8	93.4	

Table 6. 1: Diagnostic accuracy (%) as a function of training and fault type

6.2.2.2 *Secondary task performance*

Annunciator acknowledgement (reaction time)

This measure indicates the time needed (in secs) to respond to the appearance of an annunciator. The data for this measure are presented in Figure 6.3. Whereas no effect of training or noise was found (all F<1), the analysis revealed that the more difficult the system faults became, the more the response time slowed down. This effect was highly significant (F=16.18; df=3,69; p<.001). Post-hoc LSD-tests confirmed that the differences between all four levels of fault type were significantly different from each other (at p<.05 level or higher). No interactions were observed.

Prospective memory (tank level recording)

Recordings were assigned to four different categories according to the accuracy of timing: *Early* (5 sec or more before scheduled time), *On-time* (within 5 sec of scheduled time), *Late* (between 5 to 20 sec after scheduled time) and *Omissions* (missed or > 20 sec after scheduled time). The occurrence of early responses was generally very low (< 1%). Most recordings were made on time (ca. 45%), followed by late responses (ca. 38%) and a not inconsiderable number of omissions (ca. 16%). The category of omissions emerged as the most sensitive one and was therefore selected for the analysis. The data are presented in Figure 6.4.

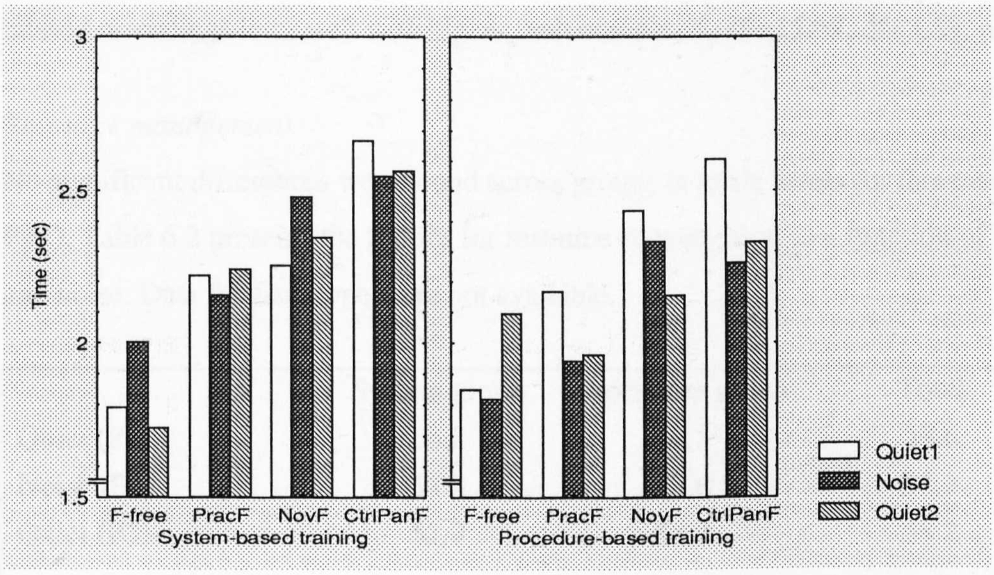


Figure 6. 3: Annunciator response time (sec) as a function of training, noise and fault type

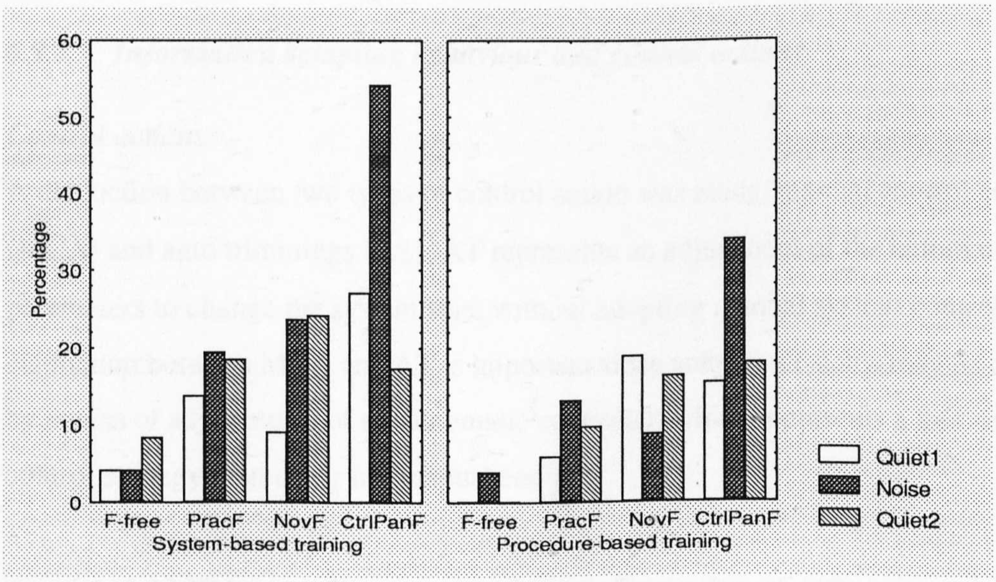


Figure 6. 4: Omitted tank level recordings (%) as a function of training, noise and fault type

The results showed that the system group missed more recordings than the procedure group (18.7% vs 12.0%) but the result was not significant (F=2.04; df=1,23; p>.05). Furthermore, the results showed that participants more frequently missed tank level recordings when exposed to noise (Q1: 11.8%, N: 20.1%; Q2: 14.0%). This difference was significant (F=5.08; df=2,46; p<.05). Prospective memory was particularly vulnerable under noise when combined with the most difficult fault condition (see Figure 6.4). This resulted in a more than 100% increase in the number of omissions for both groups (F=3.77; df=6,138; p<.005). Finally, the analysis revealed that the number of missed recordings increased with higher fault complexity (F=15.05; df=3,69; p<.001).

Resource management

No significant differences were found across groups or noise levels for this measure (all F<1). Table 6.2 presents the figures for resource consumption as a function of training and noise. Data for fault type were not available.

	System group	Procedure group	Total
Quiet (P1)	56.8	56.5	56.6
Noise (P2)	58.8	57.6	58.2
Quiet (P3)	57.9	58.4	58.2
Session	57.8	57.5	

Table 6. 2: Resource consumption (%) as a function of training and noise

6.2.2.3 Information sampling behaviour and control actions

Control actions

A distinction between two types of control action was made: manual control actions (MCA) and auto trimmings (AT). AT represents an adjustment of the automatic control parameters to change the system state without adopting manual system control. The distinction between MCA and AT is important since some fault states may be controlled by means of adjustments of the automatic controller, which represents a less demanding control strategy than using full manual control.

The data for MCA and AT are presented in one figure (Figure 6.5), averaged across periods, since no effect of noise was found for either measure. This was unexpected since one may presume that more ATs are used during noise exposure because of this control mode requiring fewer mental resources. The faultfree condition was not included in this analysis since virtually no control actions were carried out during these phases. First, the results for auto trimmings are examined. The analysis of variance did not reveal any significant effects although ATs were more frequent during NovF than during the other two fault states (F=1.86; df=2,46; p>.05). As Figure 6.5 indicates, no differences were found for training (F<1). An effect of noise was not observed (F=1.36; df=2,46; p>.05), though a slight reduction in ATs was recorded in the last hour. The analysis of MCA indicated that system-trained participants carried out nearly twice as many control actions than their procedure-trained colleagues (0.63 / min vs 0.36 /min). The main effect was not significant (F=3.41; df=1,23; p>.05), but examining the *training x fault type* interaction showed a very strong increase in control activity for the system group during the most complex fault condition. Figure 6.5 shows that MCA during the CtrlPanF increased much more strongly for the system group than for the procedure group, resulting in a significant interaction (F=5.91; df=2,46; p<.01). In addition, the results showed a general rise in manual control activity with augmenting fault difficulty. This was confirmed by a highly significant result for fault type (F=44.06; df=2,46; p<.001). However, post-hoc LSD-tests revealed that only the CtrlPanF conditions was significantly different from the two others (both LSD-tests: p<.001). As can also be seen in Figure 6.5, participants generally preferred AT to MCA for system management during PracF and NovF. However, this strategy appeared to be ill-suited to deal with the more difficult CtrlPanFs, resulting in a strong increase of MCA in this condition.

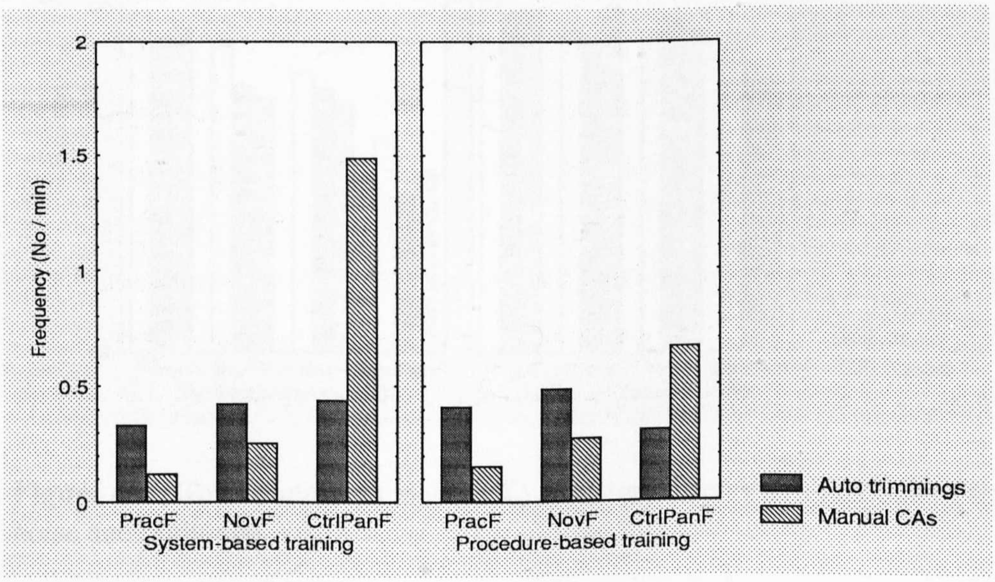


Figure 6. 5: Frequency of control actions (No/min) as a function of training, control strategy and fault type

History display sampling

Analysis of the information sampling behaviour revealed that on average participants made extensive use of the history display, which was displayed on screen for 85.5% of the time. When evoked, the history display remains visible for a period of 30 sec. The procedure group showed higher sampling times than the system group (88.0% vs 82.9%). The difference was statistically significant (F=10.86; df=1,23; p<.005). The gap between the training groups tended to widen with increasing fault difficulty (see Figure 6.6), resulting in a significant interaction effect (F=3.23; df=3,69; p<.05). As shown in Figure 6.6, higher sampling times were also found in the first hour of the experiment compared to the noise and post-noise period (F=3.51; df=2,46; p<.05). The third main effect also emerged to be significant with sampling frequencies becoming smaller with increasing fault difficulty (albeit no difference was found between PracF and NovF). In the faultfree phase, sampling times peaked at 89.3% whereas in the CtrlPanF condition, only for 80.6% of the time the history display was displayed (PracF: 86.1%, NovF: 85.7%). The effect was significant at the following level: F=18.0; df=3,69; p<.001. No other interaction effects were found.

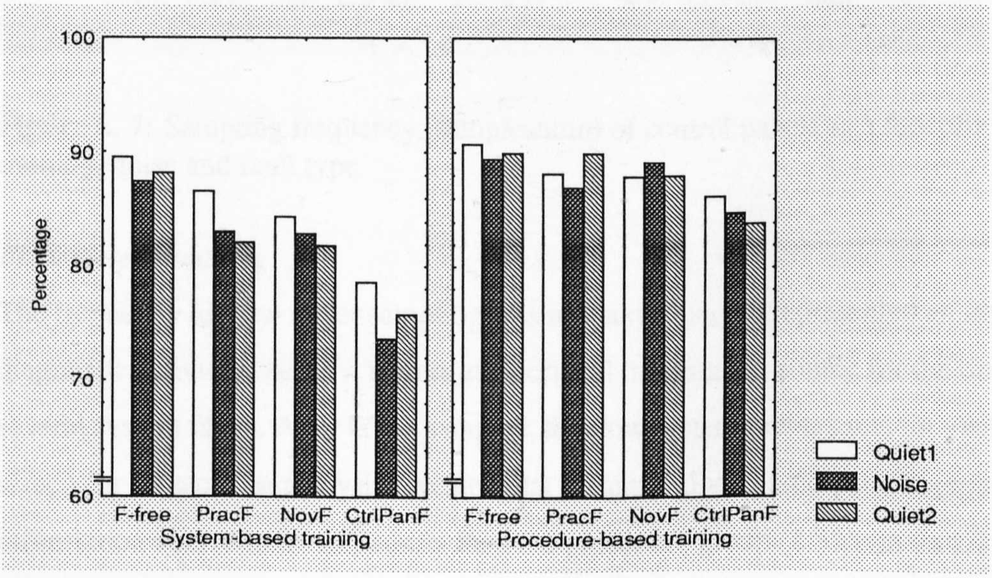


Figure 6. 6: Total sampling times (in %) of history display as a function of training, noise and fault type

Control panel sampling

An analysis of the frequency of control panel samples (excluding control actions) revealed that the system group used this information source more often than the procedure group (1.04 samples/min vs 0.83 samples/min). However, this difference was statistically not significant (F<1). The control panels also seemed to be more frequently accessed with increasing duration of the experiment (F=3.46; df=2,46; p<.05). The data for the different experimental conditions are presented in Figure 6.7. Whereas no differences were observed in this measure between F-free, PracF and NovF, the presence of a control panel failure led to a strong increase in sampling activities (F=67.6; df=3,69; p<.001). No significant interactions were found.

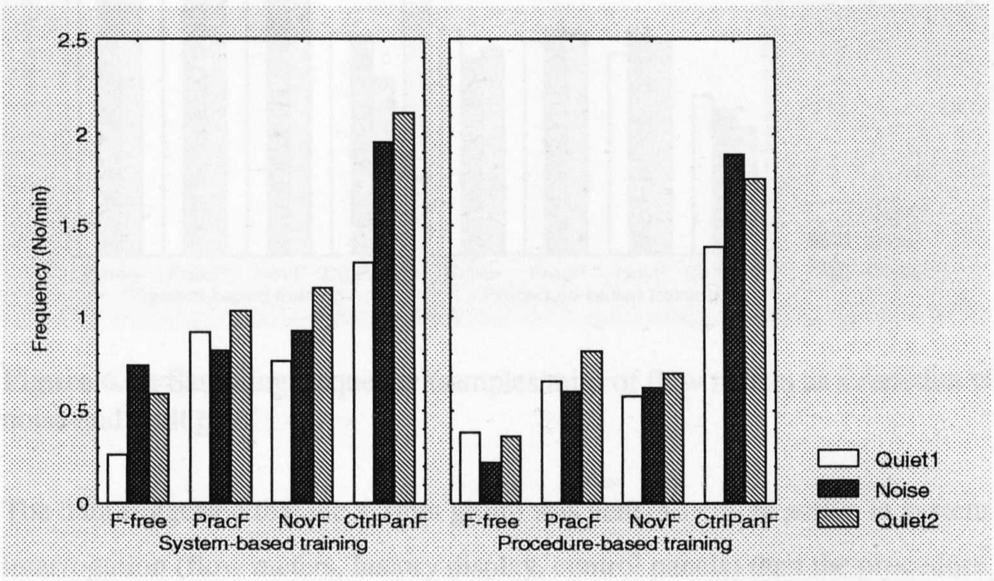


Figure 6. 7: Sampling frequency (samples/min) of control panels as a function of training, noise and fault type

Flow meter sampling

The frequency of flow meter sampling was measured since it represented an indicator of diagnostic activity. The flow meters are a critical information source for the diagnosis of several system faults. After being sampled, the flow meter readings remain visible on screen for 10 sec. The analysis revealed that system-trained participants sampled this information source more frequently than the procedure group (2.35/min vs 1.71/min). This effect was found to be significant (F=4.44; df=1,23; p<.05). The highest number of system interrogations was recorded for the PracF (2.29/min) and NovF (2.22/min) conditions whereas during the faultfree phase significantly fewer checks occurred (2.07/min). The lowest number of readings was observed for the CtrlPanF (1.53 s/min). The effect reached overall statistical significance (F=23.9; df=3,69; p<.001), and post-hoc LSD test revealed that all four conditions were significantly different from each other at the p<.05 level (except between PracF and NovF). No effect of noise (F=1.24; df=2,46; p>.05) was observed and no significant interaction was found. The data for all conditions are presented in Figure 6.8.

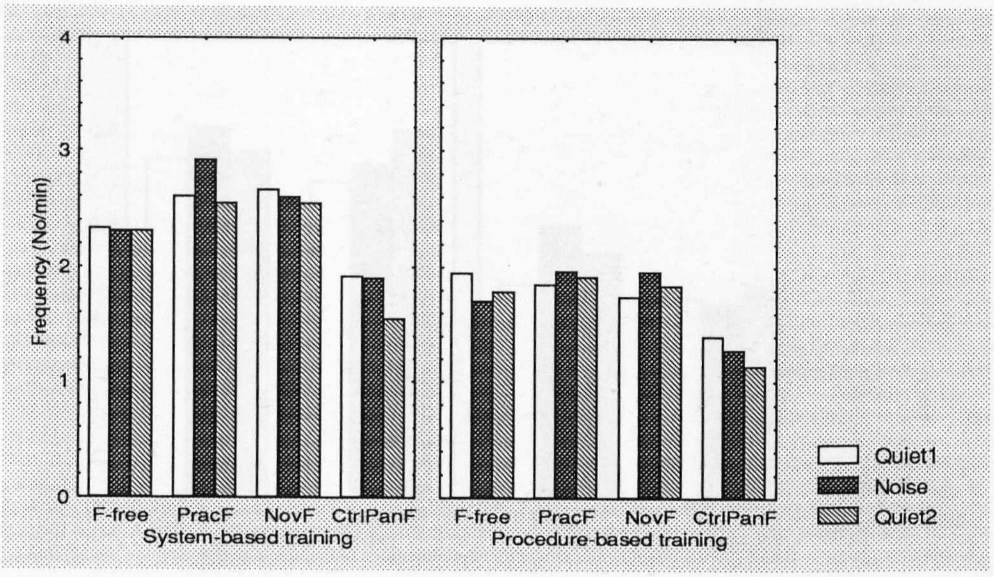


Figure 6. 8: Sampling frequency (samples/min) of flow meters as a function of training, noise and fault type

The data suggested that the system group was generally more active in system interrogation (flow meters, history display, control panels) than the procedure group. Table 6.3 summarises the sampling frequency of the different interrogative facilities of the system.

	System group	Procedure group
History display	1.66	1.76
Control panels	0.75	0.53
Flow meters	2.35	1.71
Total	4.76	4.00

Table 6. 3: Sampling frequency of different system components (samples/min) as a function of training

Interrogation of maintenance facility

Analysis of the experimental records showed that participants frequently accessed the maintenance facility without carrying out a repair action. A closer analysis of this behaviour revealed that this was more frequent for the system group (0.38 samples/min) than for the procedure group (0.25 samples/min). This difference was statistically significant (F=4.89; df=1,23; p<.05). No other main effects or any interaction was found (all F<1). The data are presented in Figure 6.9.

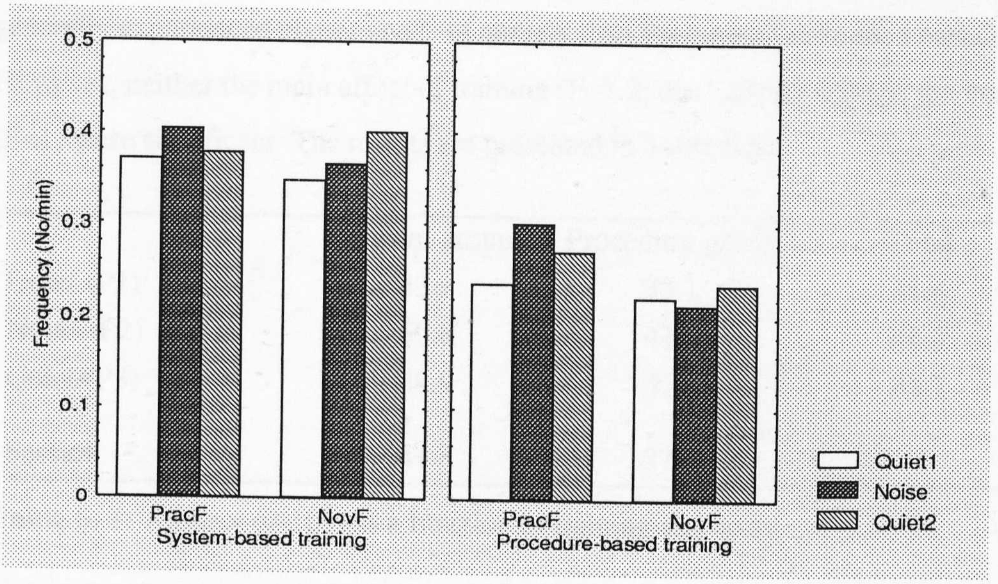


Figure 6. 9: Sampling frequency (samples / min) of maintenance facility as a function of training, noise and fault type

6.2.2.4 Subjective state

The subjective state of the participants was measured at the end of each 30-min trial and was therefore not available as a function of fault state. Responses were made on 100-mm visual analogue scales for three variables: Effort, Anxiety and Fatigue.

Effort

Task-related effort showed only minor variations as a function of the experimental conditions although effort levels were somewhat increased during noise exposure for the procedure group (see Table 6.4). However, none of the effects were significant (all F<1).

	System group	Procedure group	Total
Quiet (P1)	52.5	52.8	52.6
Noise (P2)	52.5	58.2	55.5
Quiet (P3)	52.0	53.0	52.5
Session	52.3	54.6	

Table 6. 4: Mental effort (0-100) as a function of training and noise

Anxiety

The analysis revealed a clear effect of noise on anxiety levels (F=3.97; df=2,46; p<.05), manifesting itself in an increase of anxiety during noise exposure compared to the two quiet periods. The effect appeared to be more pronounced for the system group, who generally experienced higher levels of anxiety than the procedure-trained participants. However, neither the main effect of training (F=1.2; df=1,23; p>.05) nor the interaction (F<1) were significant. The results are presented in Table 6.5.

	System group	Procedure group	Total
Quiet (P1)	43.9	39.1	41.4
Noise (P2)	50.8	41.9	46.2
Quiet (P3)	39.9	31.8	35.7
Session	44.8	37.6	

Table 6. 5: Anxiety (0-100) as a function of training and noise

Fatigue

Fatigue levels rose strongly during exposure to noise (see Table 6.6) and did not return to previous levels in the post-noise period (F=8.22; df=2,46; p<.001). This effect was more pronounced for the procedure group, who even experienced a further increase of fatigue during the last period. However, the effect was not significant (F=2.92; df=1,23; p>.05). No main effect of training was observed (F<1).

	System group	Procedure group	Total
Quiet (P1)	53.3	39.7	46.2
Noise (P2)	61.8	60.5	61.1
Quiet (P3)	59.3	67.5	63.5
Session	58.1	55.9	

Table 6. 6: Fatigue (0-100) as a function of training and noise

6.2.2.5 Mental model of operator

An attempt was made to assess the operator's mental model of the system by measuring two aspects of system knowledge: *Explicit knowledge* and *Implicit knowledge*. The evaluation of the mental model was conducted in collaboration with an MSc student (Dobson 1994), who carried out the work as part of his dissertation. His data have been reanalysed for the purpose of evaluating the development of the mental model after an extended lay-off period (see section 6.3.2.5).

Explicit knowledge

In a post-experimental interview, questionnaires (see Appendix 3) and verbal protocols were employed to assess the quality of the mental model of the participants. The verbal protocol was recorded while the participant managed the system during a fault state. Verbal protocols and the questionnaire were then analysed to assess the participant's mental model. A score was derived from the participant's knowledge about three separate aspects of system functioning (corresponding to an effective control of the three CtrlPanFs). A score of one was given for each CtrlPanF, allowing each participant to achieve a maximum score of three. The criterion for this was a clear demonstration of the understanding of the system functioning relevant to this particular fault state (e.g. the inflow of liquid N_2 will reduce the cabin temperature and can therefore be used in the event of a failed cooling system).

Implicit knowledge

Implicit knowledge was measured by looking at participant performance during the occurrence of the control panel failures and by asking the participants to give a brief demonstration of their control strategies on screen during the post-experimental interview session. Likewise, a maximum score of three could be obtained if participants used the correct control strategies for all three faults. The criterion for receiving a score of one was a clear on-screen demonstration of the control strategy (or in some cases one of several) that was appropriate to deal effectively with the control panel failure.

The post-experimental interview sessions were carried out by Dobson (1994), who then rated the knowledge of the participants on the basis of the data sources described above. Using the same data sources, the evaluation of the mental model was carried out again for this thesis to obtain an independent rating, permitting the computation of an inter-rater reliability coefficient.

Inter-rater reliability

The results suggested that there was a considerable degree of agreement between raters concerning the participants' system knowledge. The inter-rater reliability coefficient for all fault states together was $r = .85$ ($df=23$; $p<.001$). A separate analysis of the implicit and explicit knowledge of each fault state corroborated the overall trend. Some correlation coefficients reached a value of 1 whereas the lowest correlation coefficient computed was .46 (df=23 ; p<.05) for implicit knowledge of control panel failure 1. The level of inter-rater reliability was equally high for explicit (r=.89) as for implicit knowledge (r=.87).

Knowledge and training

The results suggested that system-trained participants had a higher system knowledge than their colleagues trained on procedures (F=10.15; df=1,23; p<.005). Table 6.7 indicates that the system group showed superior knowledge for both knowledge types. Although explicit knowledge showed a higher score than implicit knowledge, the difference was statistically not significant (F<1).

	System group	Procedure group
Explicit knowledge	1.41	0.69
Implicit knowledge	1.25	0.46
Total system knowledge	2.66	1.15

Table 6. 7: System knowledge for different training approaches

Association between explicit and implicit knowledge

It was expected that participants showed a high level of coherence between their implicit and explicit knowledge, that is, if a participant demonstrated explicit knowledge of a particular fault, he/she should also have the implicit knowledge of that fault and vice versa. However, a considerable number of participants showed some dissociation between the two types of knowledge (see Table 6.8). On 11 occasions participants demonstrated the ability to manage effectively a given control panel failure but they were unable to explain their knowledge when interviewed. Conversely, although participants could explain how to cope with a certain control panel failure on 15 occasions, they did not demonstrate that knowledge in a practice session. This equals a total number of 26 cases (34.7%) of dissociated knowledge.

		Explicit knowledge	
		Yes	No
Implicit	Yes	11	11
knowledge	No	15	38

Table 6. 8: Association between explicit and implicit knowledge (the total number equals 75 with each participant tested on 3 faults)

6.2.3 Discussion

6.2.3.1 Training

In contrast to expectations, the training variable had overall only a limited impact on the different performance measures. It was predicted that the procedure group would be better at managing practised faults but that novel faults (as well as control panel failures) would be dealt with most effectively by the system-trained operators. This was based on the assumption that system-based training would lead to the development of a superior mental model of the life support system, which would enable the operator to better understand unfamiliar fault states and to deal with them more effectively. The results for parameter control failures showed a pattern into the expected direction (without reaching statistical significance), with the system group performing better on both unfamiliar fault states. There was also evidence that the system group had a better mental model on both dimensions (explicit and implicit knowledge). However, it is not clear why the superior mental model did not pay off in terms of improved performance. In the present case one may suspect that although the system group did better on the parameter control failures, the fault condition was of such great difficulty that even the most skilled participants had problems to complete a trial without committing any errors. Therefore, the control failure measure might not have been sufficiently sensitive to distinguish between the different strategies to control the fault state. One may even suspect that correct control strategies led to more errors if the participant had the right idea of using another subsystem to deal with the disturbance, but as a result of problems to implement the control strategy successfully, the operation caused the corresponding parameter to drift out of range. The absence of an interaction *training x fault type* is however not in conflict with findings from the literature. It has been suggested that better knowledge does not necessarily lead to superior trouble-shooting performance (Johnson and Rouse 1982), and that the teaching of diagnostic rules often resulted in better performance than theoretical knowledge of the system (Shepherd et al 1977).

An unexpected finding was that system-trained participants did not show better performance on the diagnostic task for speed and accuracy when confronted with novel faults. The better mental model of the system group was expected to result in superior diagnostic performance on the previously unencountered fault states. It might be possible that the system group used a fundamentally different approach to fault identification, which consisted of a more cautious and detailed assessment before committing themselves to any decision. One may argue that the larger number of maintenance facility checks (where no repair was carried out) for the system group is indicative of the more cautious approach. The system group made more incorrect diagnoses than the procedure group, suggesting that the reason for their longer fault identification times is partly due to a larger number of failed repairs. There has been evidence for a dissociation between system control and system diagnosis (Kessel and Wickens 1982; Landeweerd 1979), with operators who are good at one are not necessarily good at the other. Although this has been attributed to different underlying abilities rather than training methods, it is conceivable that the two training regimes resulted in the development of different fault finding skills. For example, the procedure group may have made more use of the fault finding guide (external memory) than the system group who may have relied more on their system knowledge (internal memory). Unfortunately, no data are available to examine this hypothesis. In hindsight, it would have been better to collect data on the use of the fault finding guide, for example, by providing a task-embedded version of it.

However, despite the absence of marked effects of training on performance, considerable differences in information sampling behaviour were observed. This suggests that the information management strategies were substantially shaped by the training methods. The general picture was that the system group showed a propensity to interrogate the system more often than the procedure group. Furthermore, the system group showed a marked increase in manual interventions during the CtrlPanF condition, indicating a much more active approach in dealing with unfamiliar situations. This was in contrast to the procedure group which may have displayed more of a ‘laissez-faire’ attitude. The general tendency of system-trained participants to adopt manual control more frequently may be due to their superior mental model, since this is a basic requirement for successful open-loop control (Wickens 1992). System-based training may have also led to operators having more confidence in their skills since previous research has indicated that more confident operators make more use of manual control skills (Lee and Moray 1994).

6.2.3.2 Noise

Noise as the second independent variable showed no significant effect on primary task performance. However, a trend in the predicted direction was observed, with performance under noise being impaired for the system group whereas no impairment was recorded for the procedure group. This (non-significant) trend was only observable for the more demanding fault states (NovF and CtrlPanF). It was assumed that the training methods would cause the system group to adopt a more knowledge-based system management approach (in Rasmussen's terms), as opposed to the procedure group who was expected to use a more procedure-based strategy. The higher level cognitive activity of knowledge-based control was supposed to be more vulnerable to noise. The fact that no significant effects were found indicated that the participants managed to avoid decrements in primary performance. As a consequence of this, the negative impact of noise seemed to manifest itself in the prospective memory task. This measure showed a strong decrement but only under the most difficult fault type. With the percentage of omitted recordings approximately doubled, it indicated that the more demanding activities such as real-time problem-solving during CtrlPanFs had excessive resource requirements and therefore could not accommodate an additional stressor such as noise.

The research literature on noise has provided a complex picture of effects with some tasks showing improvements under noise while others suffer (Holding 1983; Jones 1983; Hockey 1986). However, it suggests that for those tasks which are strongly dependent on working memory, some form of impairment is to be expected. Since the prospective memory task was considered to require the highest working memory resources among the secondary tasks, it confirmed the expectation that it would suffer first under noise exposure. In contrast, noise did not show any effect on performance on the annunciator acknowledgement task because this activity was thought to require less working memory capacity and was therefore less vulnerable.

Even though performance was largely unaffected by noise, some changes in information sampling behaviour were observed with operators turning more towards the control panels for system monitoring while the history display was less frequently sampled. It is not entirely clear whether this resulted from noise exposure or whether it was simply a function of time-on-task (or a combination of both), since the third (noise-free) period did not show a return to pre-noise levels. This study did not confirm the results found by Dörner and Pfeifer (1993), where noise exposure was associated with an increase in the overall number of control actions. The information sampling data did not provide any evidence for attentional narrowing, as was observed in previous studies in the form of neglecting peripheral displays during noise exposure (e.g., Hockey 1970a, 1970b).

Of great interest was that noise exposure had a considerable effect on anxiety and fatigue, resulting in a strong increase. As for fatigue, levels did not recover during the third period, an effect that was even more pronounced for the procedure group. This could be a direct result of noise, as found in previous studies (Cohen and Spacapan 1978). However, it is more likely that this was the result of an extended working session with noise acting as an amplifier because the increase during the second hour was larger than during the third hour. Although the research literature suggested a number of performance-related changes (such as lapses in attention, decrements in peripheral tasks) as a result of increasing fatigue (Holding 1983), in the present experiment performance was largely maintained over the full duration of the session. This was also found in a recent simulation study, using a continuous 4-hour working session, in the operational context of ship's bridge control (Sauer, Crawshaw, Hockey and Wastell 1996). With increasing watch duration, participants showed a significant rise in fatigue levels but no performance impairment was observed.

Anxiety, on the other hand, returned to previous levels, following exposure to noise. It has been suggested that anxiety is caused by failures to meet implied or expected standards (Davis 1948). The feeling of not meeting the expected standard is based on individual perceptions, and noise may cause individuals to underestimate their performance. However, no evidence was found in the literature to support or refute this assumption (Holding 1983). Even in the absence of a satisfactory explanation of the underlying processes, the results provided sufficient evidence for increased stress levels under noise, which may result in fatigue after-effects or in other manifestations of work-related stress.

Mental effort as the third subjective state variable did not show any effect of noise. This is supported by evidence in the literature showing that fatigued participants are reluctant to increase their task-related effort (Holding 1983). Instead, they employ low-effort strategies during the experimental session, such as to take increased risks during decision-making (Holding 1983; Shingledecker and Holding 1974). However, no direct evidence of increased risk-taking on the diagnostic task was observed in this experiment since the number of misdiagnoses remained stable throughout the session.

6.2.3.3 Fault difficulty

This independent variable showed an effect on all performance measures in the expected direction. A strong effect on parameter control performance is not a surprising result since the fault states had a direct impact on parameter levels, leading almost inevitably to control errors. Even for the most skilled participants, it was virtually impossible to complete a trial without committing control errors since certain fault conditions were very difficulty to manage. However, secondary task performance was also affected, which indicated that the fault states required a strong commitment of cognitive resources. Even for reaction time, the least vulnerable task among the secondary task measures, a clear rise in response time with increasing primary task load was observed.

Increased fault difficulty (particularly the CtrlPanF condition) led to a general shift from monitoring behaviour (reduced viewing of history display and flow meters) to more control activity. The only exception to this was the sampling of the control panels. This increased strongly under CtrlPanF, probably because participants could use the same facility for two purposes: interrogation and intervention. This is to be considered an effective strategy since it reduced the monitoring load. Furthermore, the potential advantage of the history display, which is the identification of system faults by their unique graphical pattern, did not come into effect during this fault condition. Changes in strategies were always more pronounced (albeit not always statistically significant) for the system-trained participants, which may be considered as support for the argument that their training regime had equipped them with a more flexible and holistic approach to system management.

6.2.3.4 Mental model

The different training methods employed in the experiment resulted in the system group developing a higher level of implicit and explicit knowledge of the system. This was in line with the rationale behind the training procedures, which assumed that the focus of system-based training on the underlying principles of the system would result in a better mental model. The higher implicit knowledge score of the system-trained participants indicated that they had a better understanding of the strategies to be used for CtrlPanFs.

However, this was not reflected in the primary performance measures. The reason for this might have been that the control failure measure was not sufficiently sensitive to discriminate between poor and good control strategies since, even during the use of appropriate control strategies, primary parameters could often not be contained in the target state.

Implicit and explicit knowledge appeared to be rather loosely associated, with about a third of the participants showing some form of inconsistency between theoretical knowledge and demonstrated ability. Broadbent, Fitzgerald and Broadbent (1986) found that practice may improve implicit knowledge without having any effect on explicit knowledge. Conversely, it was found that good instructions can improve explicit knowledge without leading to an improvement in implicit knowledge (Berry and Broadbent 1984). These findings have some implications for the evaluation of training. If one only assessed the competence of an operator by an oral or written test (explicit knowledge), one may make an incorrect judgement about the actual ability of the operator to carry out the task (implicit knowledge). Conversely, those operators who have demonstrated knowledge in practice may have done so on the basis of some erroneous beliefs about the system.

6.2.3.5 Implications

The results of this study indicated that different training regimes may not show direct benefits or disadvantages when merely examining task performance. However, taking the analysis a step further revealed fundamental differences in the knowledge structure and in control and information sampling behaviour. With the difference in behaviour between training groups widening with increasing fault difficulty, one may suspect that the more difficult the fault state is, the more beneficial system-based training will become. Conversely, for faults of a lesser complexity, systems are probably better controlled by the procedure-trained operators. However, the critical level of complexity (representing the point when system-based training yields better results than procedure-based training) may be positioned a great deal closer to the upper end of the continuum than previously thought. However, some inconsistencies were observed, notably the poorer diagnostic performance demonstrated by the system group. This underlines the dissociation between system control and diagnosis as representing different abilities, which needs to be addressed in the design of training.

Noise as an environmental stressor manifested its effects on performance only on the most vulnerable task component. Effects on subjective operator state were also observed together with changes in information sampling behaviour. This suggests that operators maintained their overall performance levels but probably at a cost, of which the consequences may become manifest only during a considerably longer testing period. Increased task difficulty did not only lead to primary performance impairment (which was expected) but also resulted in secondary task performance decrements and marked adaptive changes in information sampling behaviour and intervention strategies.

Overall, the data of this study support the idea that examining performance alone is insufficient to obtain a satisfactory picture of system management under various working conditions. Instead an approach is required which includes different performance indicators (secondary task approach), information sampling behaviour, intervention strategies and subjective operator state. One may also add psychophysiological measures to that list since they represent excellent complementary measures to study adaptive responses to difficult working conditions.

6.3 Experiment 3

As pointed out in Chapter 1, skill retention gains in importance as the duration of space missions increases. Especially low-frequency tasks may suffer if the appropriate skills are not practised for a long time. Even though on-board computer simulations can be employed for the practice of low-frequency tasks, it still needs to be determined which skills are most at risk to achieve an optimal design of an on-board skill maintenance programme.

This experiment aimed to investigate the effects of an extended lay-off period on performance levels. Furthermore, it built on the questions addressed in Experiment 2 to determine which training method was more resistant to skill loss. The same participants from Experiment 2 completed another testing session after an interval of eight months without practice, allowing for a comparison of their scores before and after the lay-off period. Furthermore, having several cognitive tasks embedded in AMT allowed us to ascertain whether certain tasks were more at risk than others. It was expected that the system group would show better skill retention than the procedure group, since the literature suggests that procedure-based operations are difficult to remember over periods of non-use. Furthermore, in correspondence to findings from the research literature, it was hypothesised that the mental model would deteriorate as a result of the lay-off period.

6.3.1 Method

6.3.1.1 Experimental Design

A repeated measurement design was used with three independent variables: *Skill retention*, *Type of training* and *Fault difficulty*.

Type of training was a between-subject variable and was varied at two levels: System-based vs. Procedure-based (see Experiment 2). *Fault difficulty* was varied within sessions at four levels: Faultfree, practised fault, novel fault and control panel failure (see Experiment 2). The degree of *skill retention* was measured by comparing the results of two testing sessions. The first one measured immediate skill retention and was completed as part of Experiment 2. The second testing session, completed 8 months later, measured long-term skill retention. Concerning the set-up of system faults, the 1-hour testing period of this experiment corresponded exactly to the first hour of the testing session in Experiment 2.

6.3.1.2 Participants

Seventeen participants were tested in this experiment, which represented 68% of the original participant pool from Experiment 2. Ten of them had received procedure-based training whereas the other seven belonged to the system group. The average age of the participants was 25.1, ranging from 21 to 31 years.

6.3.1.3 Procedure

The experimental session lasted for approximately 2 hours. It started with a 3-minute 'warm-up' session to allow the participants to familiarise themselves with the task again. They were also reminded of all the tasks they had to complete during the experimental session and the different priorities attached to them. After the brief 'warm-up', a 1-hour testing session was completed. Furthermore, participants were presented with two short sessions (between 5 and 10 min) containing the two control panel failures that were not presented in the previous testing session. This allowed for a comparison to be made between the two testing sessions regarding the participants' ability to manage the most difficult fault type. As in all other experiments, the participants did not know which type of fault was to occur at what time. At the end they were given a questionnaire (see Appendix 4) to gain some information about their mental model of the simulation task.

6.3.1.4 Dependent variables

The dependent variables used were identical to those in Experiment 2 (see section 6.2.1.6) and are therefore only briefly summarised. The three primary task indicators measured were: *Parameter control failures, Diagnostic speed* and *Diagnostic accuracy.* Data were collected from the following three secondary task measures: *Annunciator acknowledgement (reaction time), Tank level recording (prospective memory)* and *Resource consumption.* The following system interrogation and intervention measures were collected: *Control actions, History display sampling, Control panel sampling, Flow meter sampling* and *Interrogation of maintenance facility.* The following subject state variables were measured: *Mental effort, Anxiety* and *Fatigue.* The mental model was evaluated with regard to *explicit* and *implicit knowledge.*

6.3.2 Results

To reduce the skewness of the distribution, a log-transformation was carried out for all performance variables and most of the control and information sampling measures. However, for the sake of clarity the means reported represent the original values.

6.3.2.1 Primary task performance

Parameter control failures (PCFs)

Prior to analysis, the data were subject to a log-transformation to reduce variance heterogeneity. Skill retention as the main research question of this study provided an interesting picture (see Figure 6.10). Surprisingly, participants generally showed improved performance after the lay-off period (PCF: 2.2%) compared to their first testing session (PCF: 3.9%). This effect was significant (F=5.51; df=1,15; p<.05). However, this was largely due to a strong interaction between fault type and skill retention (F=12.5; df=3,45; p<.001). As for practised faults, participants did not manage to retain their skills, with PCFs more than doubling from 0.4% to 0.9% at the second testing session (LSD-test: p<.05). The opposite trend was observed for control panel failures with participants committing only half as many PCFs at the second testing sessions (12.9% vs 6.1%; LSD-test: p<.001). No significant effect of skill retention was found for novel faults although the trend indicated some improvements. As before, a strong trend of fault type was found in the predicted direction (F=96.8; df=3,45; p<.001). No effect of training was found (F=1.23; df=1,15; p>.05) although the system group committed overall fewer PCFs than the procedure group (2.8% vs 3.4%).

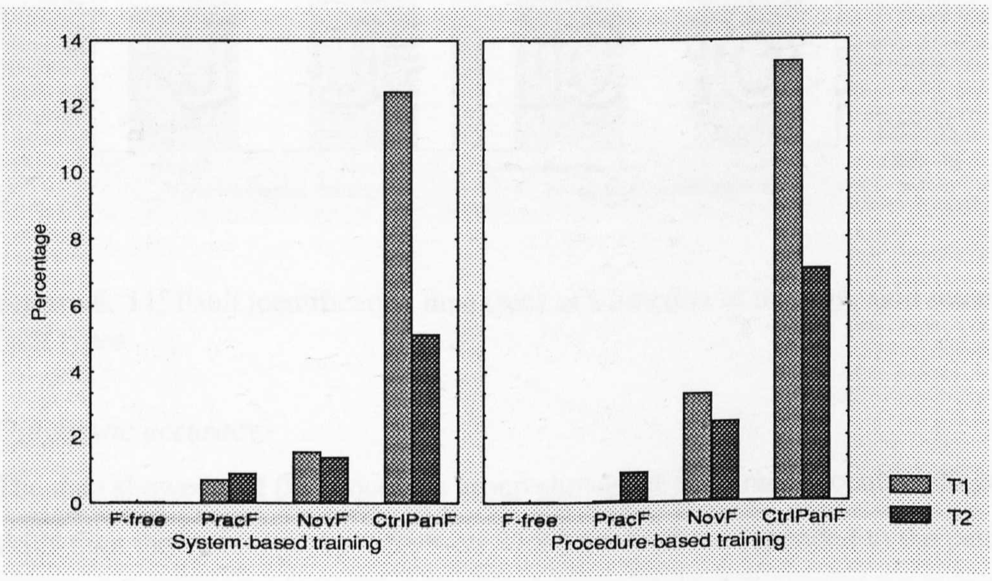


Figure 6. 10: Parameter control failures (%) as a function of training, skill retention and fault types

Diagnostic speed

The data for diagnostic speed are presented in Figure 6.11. The analysis revealed a significant rise in fault identification time at T_2 (F=6.35; df=1,15; p<.05). Participants needed on average 20% longer to make the correct diagnostic decision. This increase was mainly accounted for by the procedure group since their performance deteriorated at the second testing session (113 sec vs 131 sec) whereas little difference was found for the system group (110 sec vs 113 sec). However, the interaction failed to reach statistical significance (F=3.73; df=1,15; p>.05). In both testing sessions, fault detection times were longer for the novel than for the practised faults (145 sec vs 78 sec). This result was highly significant (F=79.8; df=1,15; p<.001). The analysis also revealed an interaction between training and fault type (across T_1 and T_2) with the system group doing better on PracFs but less well on NovFs than the procedure group (F=4.81; df=1,15; p<.05).

However, it appeared that at T_2 the difference in diagnostic speed between fault types became less pronounced for the system group while it became more pronounced for the procedure group. The 3-way interaction just failed to be significant (F=4.32; df=1,15; p=.055).

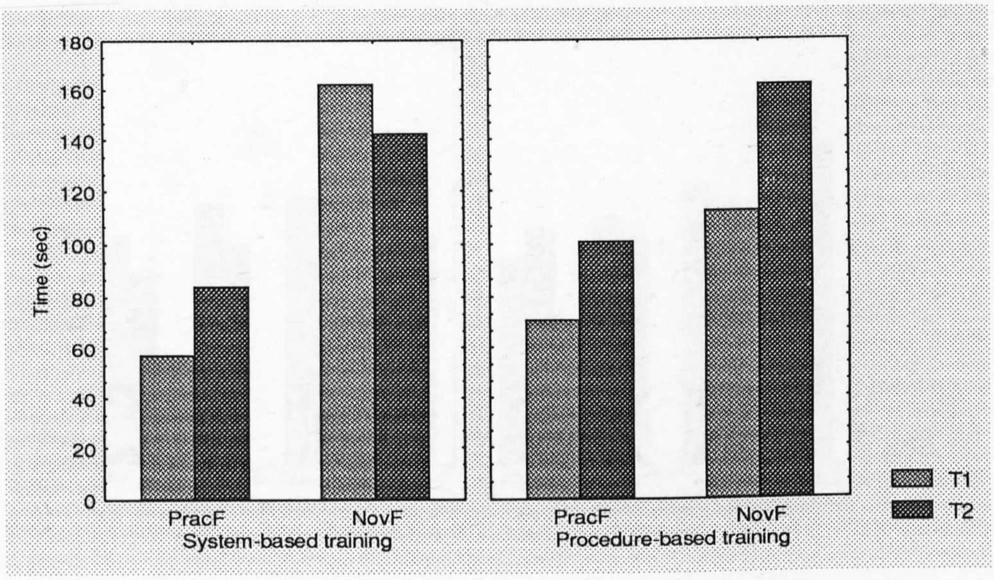


Figure 6. 11: Fault identification time (sec) as a function of training, skill retention and fault types

Diagnostic accuracy

The data showed that the procedure group showed slightly more accurate diagnostic skills than the system group (85.9% vs 82.8%), which corresponded to the pattern found at T_1. However, again the difference was not sufficiently large to gain statistical significance (F<1). Diagnostic accuracy was also found to have deteriorated slightly after the lay-off period (from 86.9% to 81.8%) but this change failed to be significant (F<1). As already observed at T_1, the NovFs were less accurately diagnosed than the PracFs (77.5% vs 91.2%). This result was significant (F=7.59; df=1,15; p<.05). No significant interaction was found (all F<1).

6.3.2.2 Secondary task performance

Annunciator reaction time

As already observed in Experiment 2, RT was strongly associated with fault difficulty in that increasing fault difficulty led to a slowing down of response time (see Figure 6.12). This effect was highly significant (F=9.13; df=3,45; p<.001). No significant effect of skill retention was found, though RT appeared to have improved at the second testing session (F=3.45; df=1,15; p>.05). The system group showed slightly faster reaction times than the procedure group but the difference was not found to be significant (F<1). No interaction effects were observed.

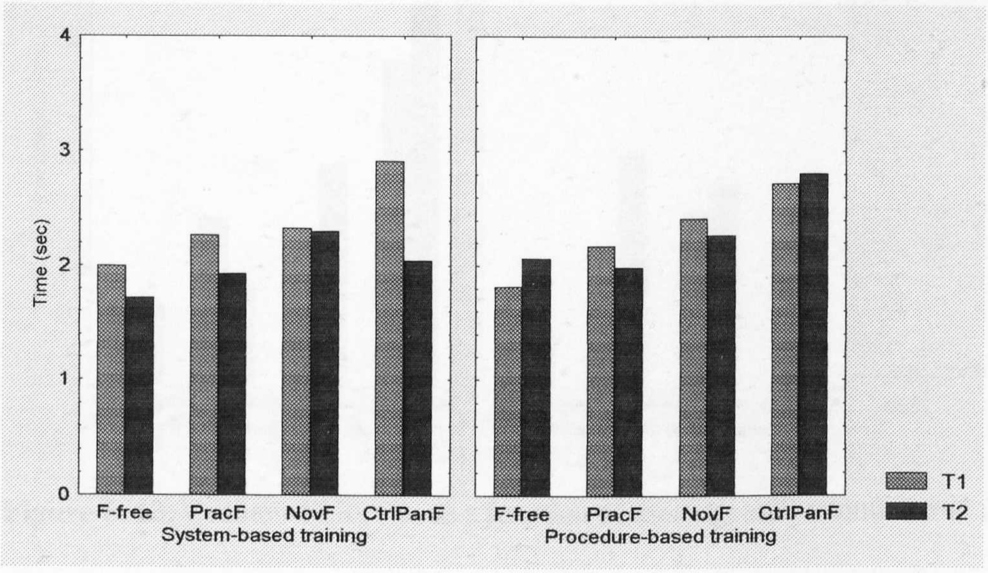


Figure 6. 12: Reaction time (sec) as a function of training, skill retention and fault type

Prospective memory (TLR)

Figure 6.13 shows the percentage of missed TLRs for the different experimental conditions. As found in most other tasks, fault type had an impact on performance with more TLRs being missed when fault states became more complex (F=3.26; df=3,45; p<.05). In addition to this main effect, an interaction between *training x fault type* indicated a more complex picture. Procedure-trained participants showed a lower percentage of omissions during CtrlPanF (8.8 %, averaged across T_1 and T_2) compared to the two less demanding fault states (PracF: 13.4%; NovF: 16.7%). This was in contrast to the system group, which followed the expected pattern of increasing error scores with higher fault difficulty (PracF: 11.8%, NovF: 12.2%, CtrlPanF: 31.0 %). This interaction was significant (F=3.02; df=3,45; p<.05). The data indicated an overall increase in omissions after the lay-off period but the difference failed to reach statistical significance (F= 3.45; df=1,15; p>.05). No other interaction was found to be significant.

Resource management

Analysis of the resource management data did not reveal any significant difference between the experimental conditions (all F<1). Table 6.9 contains the percentage of resource consumption.

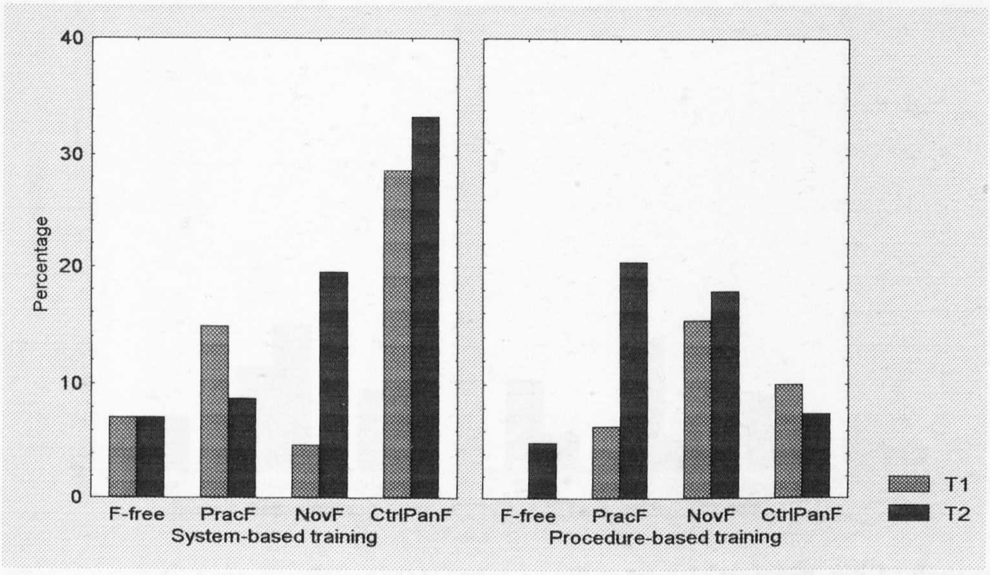


Figure 6. 13: Omitted TLRs (%) as a function of training, skill retention and fault type

T_1		T_2		T_1	T_2
SysG	ProcG	SysG	ProcG	All	All
55.8	55.9	56.0	57.5	55.9	56.7

Table 6. 9: Resource consumption (%) at testing times T_1 and T_2 for all fault states

6.3.2.3 System interrogations and interventions

Control actions

As in Experiment 2, a distinction was made between AT (auto trimmings) and MCA (manual control action). As Figure 6.14 displays, ATs were generally more frequent for NovFs than for the two other fault types. Although this pattern was not found for the ProcG at T_2, the overall effect was significant (F=3.51; df=2,30; p<.05). Post-hoc LSD-tests revealed that only the difference between PracFs and NovFs was significant (p<.05). Overall, a reduction in AT was found after the lay-off period but the effect failed to reach significance levels (F=2.10; df=1,15; p>.05). This was because the observed pattern was distinctly different for the two groups. Whereas the ProcG showed a reduction in ATs at T_2, the SysG increased the use of this control mode. This interaction was significant (F=7.28; df=1,15; p<.05). No main effect of training group was found (F<1).

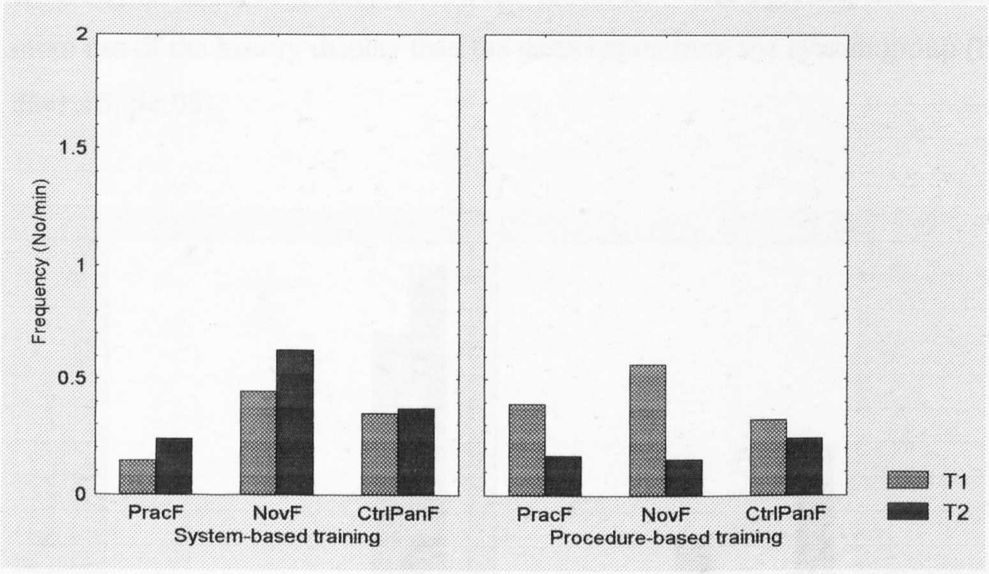


Figure 6. 14: Auto trimmings (No/min) as a function of training, skill retention and fault type

The results for MCA confirmed previous findings from Experiment 2, with an effect of fault type and an interaction *training x fault type* occurring again. The data are presented in Figure 6.15. Most MCAs were carried out during CtrlPanFs, followed by NovFs and PracFs (F=25.22; df=2,30; p<.001). The increase in MCAs for CtrlPanFs was more pronounced for the SysG than the ProcG (F=7.03; df=2,30; p<.001). A main effect of skill retention was also observed, with participants being generally more active at the second testing session. MCAs nearly doubled at T_2 compared to T_1 baseline levels, which represented a significant increase (F=14.14; df=1,15; p<.005). However, the procedure group contributed much more to this increase than the system group, which had a higher baseline level at T_1. This interaction was however not significant (F=2.59; df=1,15; p>.05). No significant difference between training group was observed (F<1).

History display sampling

The frequency of history display sampling fell from T_1 to T_2 but still remained at a high level with the graphs being on screen for most of the time (T_1: 87.5%; T_2: 83.6%). This reduction was significant (F=6.37; df=1,15; p<.05). The analysis revealed that this decrease was due to a change in sampling behaviour for the system group while the procedure group showed a virtually identical pattern (see Figure 6.16). This interaction proved to be significant (F=4.68; df=1,15; p<.05). Again, as in Experiment 2, a significant monotonic reduction in sampling duration was found with increasing fault difficulty (F=6.15; df=3,45; p<.005), and procedure-trained participants made generally more use of the history display than the participants from the system group (F=4.59; df=1,15; p<.05).

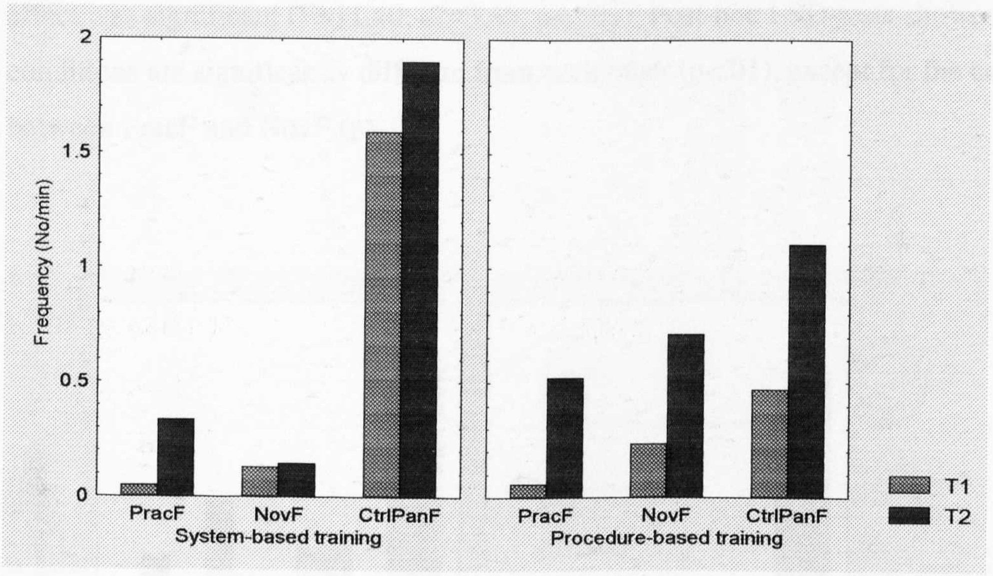


Figure 6. 15: Manual control actions (No/min) as a function of training, skill retention and fault type

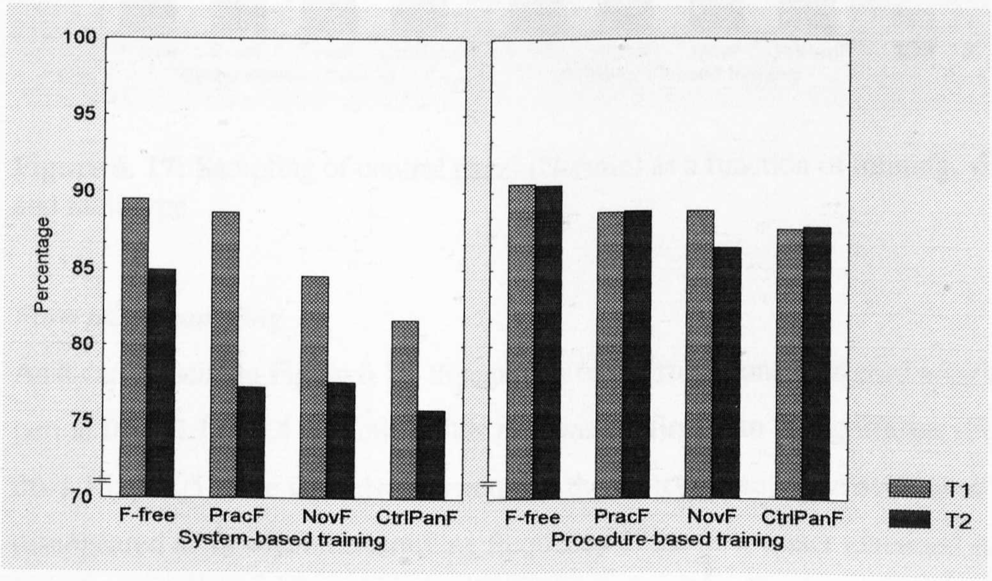


Figure 6. 16: Total duration of history display sampling (%) as a function of training, skill retention and fault type

Control panel sampling

The number of display interrogations has slightly increased after the lay-off interval (from 0.92 to 1.01 No/min) but the increase was not significant (F=1.61; df=1,15; p>.05).

Although the system group accessed the control panels more frequently than the procedure group (1.07 vs 0.87 No/min), the difference failed to be significant (F<1). As already observed in Experiment 2, participants sampled the display most frequently under the CtrlPanF condition and least frequently under F-free (see Figure 6.17). The overall effect was significant (F=11.40; df=3,45; p<.001). Post-hoc LSD-tests showed that all conditions are significantly different from each other (p<.01), except for the comparison between PracF and NovF (p>.05).

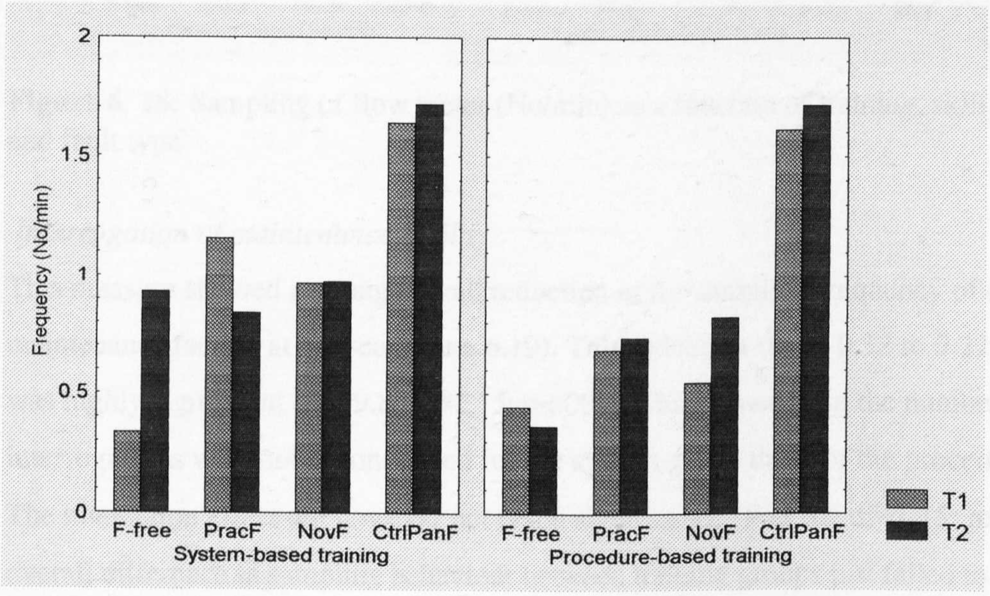


Figure 6. 17: Sampling of control panel (No/min) as a function of training, skill retention and fault type

Flow meter sampling

As it can be seen in Figure 6.18, the number of interrogations increased after the lay-off period from 2.1 to 2.4 (No/min). This rise was confirmed to be significant (F=4.74; df=1,15; p<.05). The data also showed that the effect of fault type evident at T_1 had disappeared at T_2 when the sampling frequency of the flow meter remained stable across fault type for both groups. However, this effect failed to reach significance (F=2.38; df=3,45; p>.05). Although the system group showed a higher level of sampling activity, the difference between training groups was statistically not significant (F=1.46; df=1,15; p>.05). No differences between fault types were found (F=2.01; df=3,45; p>.05). No significant interaction was observed (all F<1).

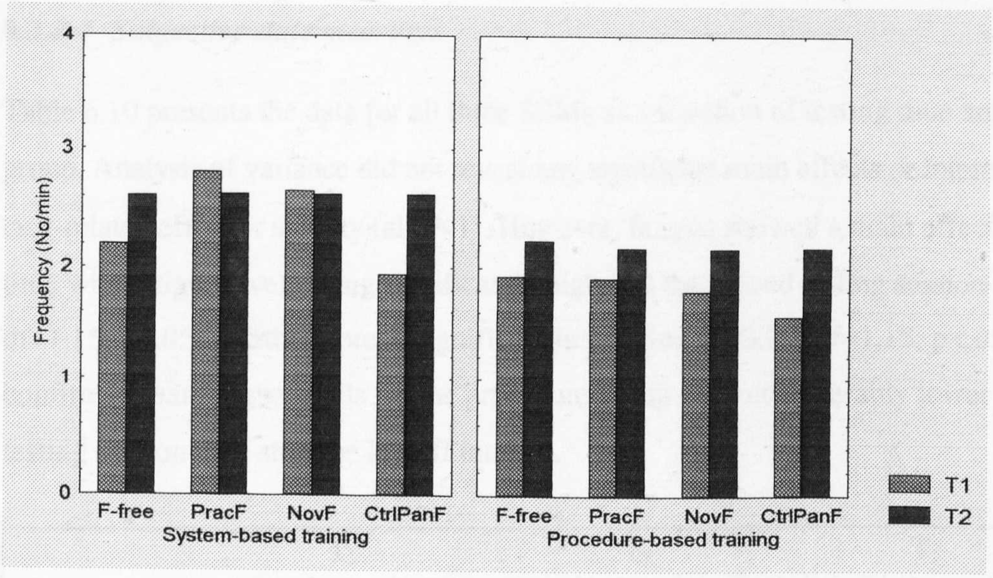


Figure 6. 18: Sampling of flow meter (No/min) as a function of training, skill retention and fault type

Interrogation of maintenance facility

This measure showed a strong overall reduction in the sampling frequency of the maintenance facility at T_2 (see Figure 6.19). This reduction (from 0.32 to 0.22 No/min) was highly significant (F=19.8; df=1,15; p<.001). This reduction in the number of interrogations was more pronounced for the system group than for the procedure group. The interaction effect did however not reach significance (F=2.62; df=1,15; p>.05). The overall difference in sampling behaviour between training groups just failed to be significant (F=4.34; df=1,15; p>.05). No other significant effects were observed.

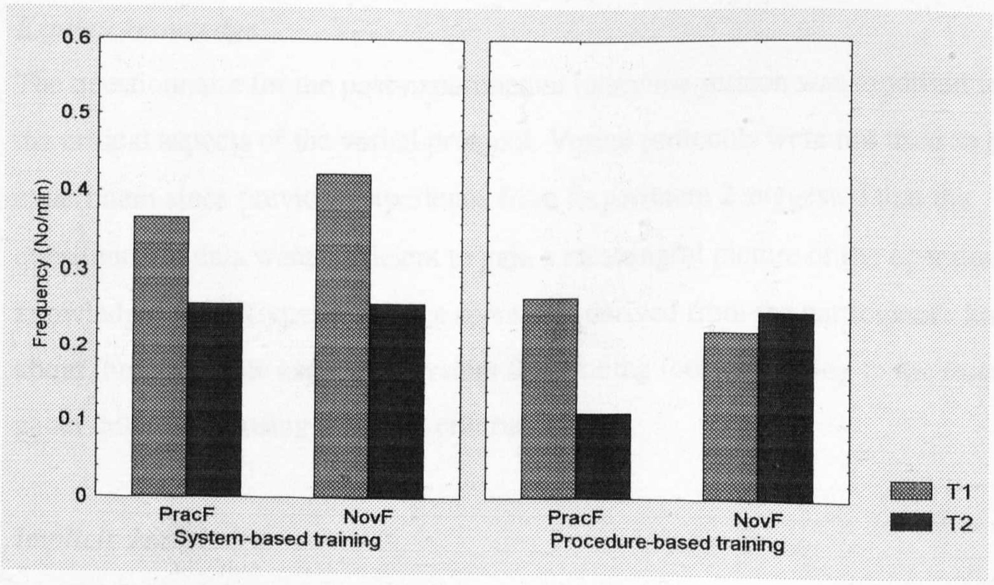


Figure 6. 19: Sampling of maintenance facility (No/min) as a function of training, skill retention and fault type

6.3.2.4 Subjective state measures

Table 6.10 presents the data for all three SSMs as a function of testing time and training group. Analysis of variance did not reveal any significant main effects or interactions for task-related effort or anxiety (all F<1). However, fatigue showed a main effect of testing time with fatigue levels being significantly higher at the second testing session (F=1.67; df=1,15; p<.05). Furthermore, a significant interaction (F=3.08; df=1,15; p<.05) confirmed that fatigue levels for the procedure group were considerably lower at the first testing session than after the lay-off interval.

	T_1		T_2		T_1	T_2
	SysG	ProcG	SysG	ProcG	All	All
Effort	48.7	53.0	51.9	53.9	50.9	52.9
Anxiety	44.7	36.5	38.2	38.6	40.6	38.4
Fatigue	53.9	39.3	51.7	53.7	46.6	52.7

Table 6. 10: Subjective state measures (0-100) at testing times T_1 and T_2

6.3.2.5 Mental model

The quality of the mental model was evaluated by using a similar method and the same criteria as described in Experiment 2 (see section 6.2.2.5). Again, two aspects of system knowledge were measured: *explicit* and *implicit* knowledge.

Explicit knowledge

The questionnaire for the post-experimental interview session was modified to include the critical aspects of the verbal protocol. Verbal protocols were not used in the present experiment since previous experience from Experiment 2 suggested that the questionnaire data were sufficient to gain a meaningful picture of the operator's implicit knowledge. As in Experiment 2, a score was derived from the participant's knowledge about three separate aspects of system functioning (corresponding to the three control panel failures) by using the same criteria.

Implicit knowledge

Implicit knowledge was measured by examining participant performance during the occurrence of the control panel failures. The demonstrated control strategies were subsequently evaluated against an optimal strategy. Again, a score of one was awarded for a successful demonstration on each CtrlPanF.

Knowledge and training

The results of the analysis are presented in Table 6.11. They indicated that system trained participants had a higher level of system knowledge at both testing times and for both types of knowledge (F=10.3; df=1,15; p<.01). It was also found that system knowledge increased over the retention interval for both training groups and type of knowledge (F=4.8; df=1,15; p<.05). No interaction was found (F<1).

	T_1		T_2		T_1	T_2
	SysG	ProcG	SysG	ProcG	All	All
Explicit knowledge	1.2	0.5	1.6	1.1	0.8	1.3
Implicit knowledge	1.4	0.2	1.7	0.7	0.8	1.2
Total system knowledge	2.6	0.7	3.3	1.8	1.6	2.5

Table 6. 11: System knowledge as function of training and skill retention

Association between explicit and implicit knowledge

As outlined above, coherence was expected between implicit and explicit knowledge, with participants either demonstrating knowledge of both types on a particular fault or failing to do so. As in Experiment 2, a considerable number of participants showed some dissociation (see Table 6.12) between the two types of knowledge (in 31.3% of cases). They demonstrated the ability to manage effectively a control panel failure in 11 cases but were unable to back it up with implicit knowledge. Conversely, on five occasions they had the theoretical knowledge about the control panel failure but failed to demonstrate that knowledge on the task. Again, no differences where found between training groups with regard to knowledge dissociation.

		Explicit knowledge	
		Yes	No
Implicit	Yes	10	5
knowledge	No	11	25

Table 6. 12: Association between explicit and implicit knowledge (Total number equals 51 with each participant tested on 3 faults)

6.3.3 Discussion

The results suggested that overall performance remained remarkably stable even after an extended period of 8 months without practice. This is an encouraging outcome since it indicates that skill retention is an issue that requires careful attention in mission planning but does not represent an insurmountable problem. The detailed analysis of performance on each task revealed a very interesting picture with secondary task performance unchanged after the lay-off period whereas control performance and diagnostic speed were affected but in different directions.

It is intriguing that control errors increased for practised faults after the lay-off period, whereas those for CtrlPanF showed the opposite trend. One would generally expect performance degradations to occur rather than improvements after an 8-month lay-off period. Models of skill decay (Rose 1989) would predict that PracFs were to suffer smaller time-induced decrements than the other conditions because considerably more practice had occurred for this fault type. Therefore, any refresher practice (such as the testing session at T_2) should be most effective for the well-developed skills in terms of the time needed to reach previous performance levels. Instead, PracF was the only fault type that showed a decrement. Since this decrement in isolation is not in contradiction to current theories of skill acquisition, we rather need to address the question of why performance deterioration was absent for NovF, and why improvements were even observed for CtrlPanF.

One may suggest several hypotheses to explain this unexpected finding. First, a simple learning process might have occurred with the participants coping better with the CtrlPanF at their second occurrence (T_2) than at their very first one (T_1), notwithstanding an 8-month interval between the sessions. This applies even more to NovF since more practice was gained during the first experiment, with participants having had additional practice on the fault in the second and third hour of the experimental session. Second, it is possible that the post-experimental session at T_1, in which the mental model of the participants was evaluated, led to knowledge improvements since it encouraged participants to think about the fault states and possible control strategies. In addition, this session may have resulted in an increase in their confidence in dealing with the fault at T_2, since they may have become aware that there was a possibility to cope with the system disturbance, although some participants may not have known the solution. Third, it cannot be excluded that participants exchanged information about control strategies after the first experiment, which may have helped them to deal with the fault in a more efficient way at T_2.

Despite the difficulties in finding a satisfactory interpretation of the results, it is encouraging that implicit system knowledge of this kind might be more resistant to decay than expected. Although the demands of the parameter control task go far beyond a simple tracking task, it still contains some fundamental elements of a tracking loop. This may explain its resistance to skill loss as previous research indicated that performance on tracking tasks showed little impairment as a result of non-use (Rose 1989; Annett 1989). The reduction in control errors is underpinned by an improvement of the participants' mental model after the lay-off interval. Although the observed improvement of the mental model was consistent with the results from the performance measures, it was at odds with findings from the research literature, which has considered mental models as temporally unstable and as being subject to a trend of simplification (Norman 1983).

The recollection of diagnostic rules appeared to have been affected by non-practice, resulting in longer identification times while accuracy was maintained. This effect was mainly accounted for by the procedure group, suggesting that their presumed strategy of remembering procedures rather than using their system knowledge may have been less resistant against time-induced perishability. This is supported by evidence in the literature, which suggested that a set of procedures is difficult to retain over time (Rose 1989; Hagman and Rose 1983). However, unlike in these studies, our participants appeared not to have left out essential steps in the procedure (since diagnostic accuracy was maintained) but required more time to complete them. Maybe accuracy was unimpaired because the procedure-trained participants used the extra time to consult the fault finding guide.

Whereas none of the secondary tasks showed any effects of skill decay, the effects of fault type identified in Experiment 2 were largely confirmed at the second testing session, providing evidence for the temporal stability of these phenomena. This suggests that the overall primary task demands have remained largely unchanged since, according to the study methodology, any increase or reduction should manifest itself in changes in secondary task performance (unless it is compensated for by a rise in task-related effort which however was not the case). However, in contrast to the stability observed for the secondary task measures, considerable changes over time were recorded for control activity and information sampling behaviour, indicating a complex picture. The most interesting result for control actions was that the procedure group showed a strong increase in manual control activity (partly at the expense of auto trimmings) and showed a more similar pattern to the system group after the retention interval. This increase in manual control actions for the procedure group might be interpreted as an increase in confidence (Lee and Moray 1994). This was in contrast to the perhaps more cautious approach to system control employed at T_1, which displayed a stronger reliance on the automatic system. This strategy might have been a direct consequence of the procedure-based training method that discouraged an explorative approach towards the system. The data seemed to suggest that the increase in control activity may have contributed to improved control performance, particularly for the CtrlPanF, since auto trimmings were often insufficient to bring the system back into its target state. There was also evidence of increased flow meter sampling activity whereas the history display sampling frequency decreased for the system group. However, it is not known why these changes occurred. They might have been changes towards optimisation of information sampling behaviour.

According to Sheridan (1972) the optimal sampling strategy depends on the trade-off between two factors: (1) the level of uncertainty about the state of an unsampled information source (a function of the event rate of that source) and (2) the cost of taking that sample (in terms of resource expenditure). However, considering the complexity of AMT, it is very difficult to determine these parameters. This is made even more difficult by the fact that the factor 'sampling costs' is not only a function of the interface but also of the operator state. That means for a fatigued operator, the costs of sampling a given display are higher than for an individual in an optimal psychophysiological state.

Among the subjective state variables, only fatigue was found to be affected by any of the independent variables. However, the interpretation of the observed increase at T_2 requires great caution since responses to the subjective state slider are based on a very individual frame of reference, of which the stability over a period of 8 months is more than doubtful. Therefore, it is preferable to err on the side of caution and ignore this significant effect because of the methodological problems associated with it.

In conclusion, there was some evidence suggesting that initial differences between procedure and system group observed at T_1 became considerably less pronounced or disappeared completely as a result of the long retention interval (e.g. diagnostic speed, PCF, flow meter readings, maintenance checks). This suggests that any initial advantage gained as a result of using a certain training method would be at risk of vanishing over time. The implications of this are discussed in the following section.

Considering the overall pattern of results, one would be inclined to conclude that there was little evidence of a threat of serious skill decrement in this task environment. However, the domain appearing to be most at risk, and therefore requiring most attention in the design of top-up training programmes, are diagnostic activities, in particular if they involve a step-by-step procedure.

6.4 General discussion

There was support for the appropriateness of the overall methodology since the different outcome measures showed a complex and diverse picture of strategies used to manage the changing task demands. A study methodology restricted to performance measures alone would have resulted in a very restricted picture of the overall pattern of results with, undoubtedly, somewhat different conclusions drawn from it. Although a comprehensive study methodology of this kind has been frequently advocated, one still finds a considerable number of research studies which focus on performance measures alone.

In relation to the conceptual framework, the results are in line with the prediction that primary task performance is largely maintained despite the presence of stressors such as noise. The fact that primary performance was affected by fault type was not unexpected in view of the extreme difficulty of some of the fault states. More important, however, is that secondary task performance was also impaired. Unlike the parameter control task, the difficulty of the secondary tasks remained stable so that the observed impairment may be interpreted as being the result of a reduction in residual capacity. This means a redirection of resources may have occurred, away from the secondary tasks to increase the available resource pool for the primary task. Since subjective state measures were not taken as a function of fault type, it is not possible to determine whether these variables were affected by the difficulty of the fault state.

Despite the fundamentally different rationale behind the two training approaches, the results did not indicate any clear superiority of one over the other. Even though there were clear indications that the system group had a better mental model and was even able to demonstrate that in control performance, the approach did not pay off in terms of achieving better performance in the measures used. On the contrary, the indications were rather that procedure-based training showed faster responses for fault diagnosis. The picture was slightly different after the retention interval with the system group showing better performance than the procedure group on the very same measure.

A number of conclusions can be drawn from this complex picture provided by the data. There is a need for the effectiveness of training methods to be evaluated over an appropriate time scale, that is, a sufficient retention interval. The training method that produces the best immediate post-training performance may not be the most effective one after an extended lay-off period. System monitoring and system control seemed to be more strongly influenced by extended lay-off periods and training than task performance. This has important implications for training. It would stress the importance of regular refresher training to ensure the maintenance of desired system management strategies, such as information sampling behaviour and intervention strategies.

Research into training for extended spaceflight needs to be particularly concerned with the maintenance of skill and knowledge during long lay-off periods. Both training approaches proved to be largely successful in limiting the number and the extent of actual performance decrements. Only very few measures showed a difference between the two methods. One may suspect that under more extreme conditions (i.e. even more severe than the CtrlPanF), one would be more likely to demonstrate the advantage of one training method over another. Therefore, had one used an even more difficult fault state, a stronger difference between the training approaches might have become apparent. As we have seen in our experiments as well as in other micro-world research (Dörner and Pfeifer 1993; Hoc and Moulin 1994), under most conditions individuals manage to maintain performance by changing their strategy. Perhaps one needs to use a ‘testing the limits’ approach (see Baltes and Willis 1982) by selecting more extreme (though in practice rare) levels of task difficulty (e.g. multiple fault states in the present case) to examine whether there is an interaction with the other independent variables. This would give us an indication of which independent variables may pose a problem under extreme conditions. At the same time, we would know that under normal operational conditions such problems would not occur.

While this chapter has dealt with skill retention after extended lay-off periods, Experiments 4-6 in Chapter 7 address the issue of skill retention in the context of isolation and confinement. Although primary task performance was largely maintained under the sub-optimal working conditions in Experiments 1 and 2, it remains to be seen whether this will also be the case under isolation and confinement. This is the central question addressed in Chapter 7.

Chapter 7:

Isolation and Confinement

7.0 Summary

Three field experiments (Experiments 4-6) were carried out to assess the impact of isolation and confinement on performance, operator state and operator strategy. Experiment 4 tested a crew of four during a short spaceflight simulation of 7 days. The results indicated that prospective memory showed a considerable decrement during peak demands whereas reaction time performance remained largely unaffected. Increased effort expenditure was used as a strategy to compensate for high task demands. Over the mission the crew showed considerable improvements in their ability to cope with periods of peak demand, whereas few changes were observed for phases of low demand. Poorer control performance in the early mission phases was due mainly to reduced operator activity, rather than the use of ineffective control actions. Overall, performance variables as well as operator state measures indicated a successful adaptation to the isolation and confinement period, though some temporary disruptions to otherwise stable performance were observed.

Experiment 5 was carried out during an 8-month over-wintering expedition in the Antarctic, with a group of 10 scientists and 6 technical support personnel. As in Experiment 4, increased effort expenditure was associated with high workload. Considerable differences between the two professional groups were found. These may be explicable by differences in ability and motivation. Increased task demands led to a widening of the performance gap between the groups. In contrast to the previous experiment, an increase in workload did not affect prospective memory and an effect on reaction time was found only for the scientists. The commonly found practice effects were considerably less pronounced in this experiment.

Experiment 6 was carried out during a simulation of a long-duration spaceflight in the Mir simulator. A crew of three was isolated for 135 days and tested three times a week for 1 hour. In addition to isolation and confinement, type of system fault (practised vs novel faults) was used as an experimental variable. The results showed few differences between practised and novel faults although it had been hypothesised that performance would initially be better on practised faults, with the difference progressively disappearing with increased practice on novel faults. There was a tendency for system monitoring activities to decline as the mission progressed, whereas tentative evidence showed the opposite trend for system interventions. Overall the results suggested that the crew largely coped with extended isolation and confinement but some indications of strain were also found.

7.1 Introduction

7.1.1 Isolation and confinement

As pointed out in Chapter 1, ICE (isolation and confinement environments) may pose a serious threat to performance maintenance during extended spaceflight and thereby endanger mission success. In a number of work environments (e.g. underwater habitats, nuclear submarines) extended isolation and confinement has been regarded as a potential threat to successful task completion (Weybrew 1990; Welham 1994). This has been attributed to a number of factors. First, ICE does usually not allow the pursuit of normal recreational activities (sport, going out, etc.) to relieve stress that may have accumulated throughout the working day. Second, considerable strain is experienced because of being in quasi-permanent contact with the other crew members during work as well as leisure activities (cabin fever). Third, the degree of stimulation of the ambient environment is very low since the features of the internal habitat hardly ever change.

Although performance maintenance has been commonly regarded as an important factor for mission success, surprisingly little systematic research has been carried out so far to look at this issue. A number of studies were carried out in the former Soviet Union to investigate the effects of ICE on different cognitive activities. They suggested that certain types of performance impairments were associated with particular mission phases. Based on their research experience in numerous simulation studies and real spaceflight, they proposed a distinction between four different in-mission phases (Gushin et al 1993). A review of the American experience in extended isolation and confinement suggested that a three-phase model might be most appropriate to describe the pattern of change the crew members go through (Kanas 1987). Both authors acknowledge that the number of distinct phases and their duration may vary between missions. Nevertheless, they both suggested that the first mission days (up to day 14) were characterised by acute adaptation problems to the new environment, usually accompanied by decreased work capabilities and increased anxiety. This is followed by an extended and stable period, with depression and boredom as the predominant psychological states. The last mission days (up to 7 days) are characterised by decreased work performance and emotional instability because of the anticipation of the imminent mission end. Mid-mission day seems to represent an important point in the psychological adaptation process because the onset of psychophysiological changes in the second phase is a function of the overall mission length. This was supported by evidence from simulation studies (Gushin et al 1993). The expected psychophysiological changes associated with the second mission half were not found in those studies that finished earlier than planned, before reaching mid-mission day.

The review of previous isolation studies (Soviet as well as American) did not provide a great deal of information about crew performance on different cognitive tasks. This is despite the fact that over the last 30 years a considerable number of isolation studies, with more than 150 participants, were carried out as part of the Soviet space programme (Gushin et al 1993). However, the studies primarily focused on medical and physiological aspects of extended isolation rather than performance related aspects of spaceflight. Those studies that addressed crew performance suggested that changes in performance were associated with certain mission phases (e.g. phase 1 and 3) even though the duration of each phase may be difficult to determine. Another review of Soviet spaceflight reported that a great deal of the performance related studies were purely anecdotal, providing evidence of a number of incidents where crew members had psychiatric episodes during the mission (Bluth and Helppie 1987).

Over the past 10 years, an important series of simulation studies has been carried out by ESA. Some of these studies looked at aspects of performance and subjective state as a function of isolation time. The first major study of this kind (ISEMSI) lasted for four weeks, during which 6 male crew members were isolated in a hyperbaric chamber (Bonting 1993). A number of experiments were carried out during that period to look at different psychological, physiological and medical issues associated with extended isolation. One experiment (Værnes, Bergan, Lindrup, Hammerborg and Warncke 1993a) found no evidence of performance decrements on any of the cognitive tasks (mental arithmetic, short-term memory, reaction time). Performance on a computerised space contaminant detection task showed overall stable performance levels for all 6 crew members (Hockey and Wiethoff 1993). Employing a task which requires intensive use of working memory, this experiment found temporary performance decrements on some mission days for most of the crew, together with a general increase in mental fatigue at weekends. Rizzolatti and Peru (1993) used a number of tasks to capture several aspects of attentional control but failed to find any effect of serious degradation during the four weeks of isolation. However, they did find a change in the attentional strategies used by the crew members, which they interpreted as a deficit of attentional capacities potentially threatening performance during periods of higher demand.

A second major study (EXEMSI) was carried out by ESA to look at the impact of isolation and confinement over a longer time period (8 weeks), including a number of performance experiments (Bonting 1996). On an auditory reaction time task crew members showed an increasing response latency for rare tones (but not for the frequent ones) shortly before the end of the isolation period (Mecklinger, Friederici and Güssow 1996). A control group completing the task at the same time did not show this pattern. Using a dual task (tracking and short-term memory), Lorenz, Lorenz and Manzey (1996) did not find any impairment of short-term memory functions throughout the isolation period, but significant decrements of tracking performance were observed for two crew members. Showing an initial deterioration after entry into the chamber, their performance recovered to pre-isolation baseline scores by the end of the first half of the mission but degraded again at the end of the isolation period. Using five tests in the form of computer games to measure different cognitive activities (see Chapter 4 for a description of the Joy task), Gushin, Efimov and Smirnova (1996) found overall no serious degradation of performance. However, some crew members showed temporary fluctuations in working memory, attention and mental arithmetic, with a steady increase in fatigue, as the mission progressed. Hockey and Sauer (1996) used a contaminant detection task, which exerted intensive demands on working memory and allowed for a trade-off between speed and accuracy. Because of progressive improvements over the mission, an exponential regression function was fitted to estimate the learning rate. Using the residuals (difference between fitted learning curve and data graph) for the analysis, the results showed some relative decrements in decision-making speed for all crew members at the end of the isolation period. The performance curve of one participant indicated even absolute decrements.

A third isolation study (HUBES[1]) has been completed only recently (Værnes et al 1995), but all results are not yet fully published. Lasting for 135 days, the study aimed to provide a generic simulation of the EuroMir-95 flight with psychological issues playing a major part in the study programme. Crew performance was examined on a number of space-related tasks. Using a simulation task involving spacecraft navigation, no general impairment of operator skill was observed, but during the simulation of extreme conditions (involving 48 hours of sleep deprivation) performance suffered (Salnitsky et al 1995). Using an adaptive tracking task, Nechaev, Panfilova, Sycora and Sholtzova (1995) found improved performance during week 7-17 of the isolation period, followed by a decline in performance levels during the last two mission weeks (this corresponds broadly to the 'final effort' period). A similar picture was found in another tracking experiment (Makhrov 1995). During later mission phases, a decline in crew performance was recorded during two mission weeks. The 'final effort' effect manifested itself in the form of impaired performance in an experiment by Efimov, Gushin and Smirnova (1995), using the Joy task (described in Chapter 4). A fatigue-sensitive decision-making task (described in Hockey and Sauer 1996) was used in another study to detect any decrements in performance or changes in information management strategies (Hockey, Sauer and Wastell 1995). No serious performance decrements were observed but occasional temporary disruptions did occur in the second mission half. Effort and fatigue levels were also found to increase in the second half of the mission.

Summarising the findings from the European space programme, it seems that no strong evidence has emerged for serious performance decrements, but temporary disruptions of performance (notably at the mission end) have been observed on a number of cognitive tasks. However, most of the tasks were based on the single-task methodology, which has a low level of external validity. Task requirements in real spaceflight are characterised by multiple and often conflicting goals.

7.1.2 Aims of experiments

As the review of the literature has revealed, past simulation studies have not addressed performance maintenance in ICE as a primary issue. The three field studies presented in this chapter would allow us to examine performance maintenance during isolation and confinement in a more systematic manner than it has been possible hitherto. Regular and frequent testing of the crew, using the simulation software, would permit the detection of more subtle changes in performance and subjective state. It was hoped that this would help us to better understand the management of multiple task demands during isolation and confinement. The three field experiments varied with regard to a number of factors: operational environment, skill level of participants, frequency of testing, number of participants, duration and severity of isolation and confinement. This allowed us to focus on different aspects of the adaptation pattern in isolation and confinement. The rationale and design of each experiment are described in the corresponding sections.

The three field experiments focused on the crew work environment. Experiment 4 took place in a mock-up of a space capsule, with fully trained astronauts as participants. In Experiment 5 the operational environment was the over-wintering camp of an Antarctic expedition. Experiment 6 was carried out in a Mir space cabin during a spaceflight simulation, using research cosmonauts.

7.2 Experiment 4

This study looked at the effects of seven days of isolation and confinement inside a hyperbaric chamber during a spaceflight simulation, which was carried out by the Canadian Space Agency. The study aimed to simulate the typical range of activities during a short space mission. It comprised a payload of 23 experiments from many different sciences such as physics, physiology, psychology and medicine. Four fully trained astronauts from the Canadian space programme were selected to take part in this ground-based simulation exercise. The study provided the possibility of testing the skill levels and measuring the subjective state of astronauts on a daily basis throughout the whole mission, using AMT 2.0 as a tool. The frequent data collection allowed us to examine short-term fluctuations of these variables during the period of acute adaptation (see section 7.1.1).

It was hypothesised that some adjustment process would occur under high control load, in the form of deteriorated secondary task performance, increased effort expenditure, or possibly a combination of both. Furthermore, it was expected that at the beginning of the in-mission phase performance would deteriorate compared to pre-mission baseline levels but improve again as the mission progressed.

7.2.1 Method

7.2.1.1 Design and treatment of data

The participants were asked to complete the experimental session (lasting 30 minutes) at the end of the working day over a nine-day period. The testing period was divided into three phases: Pre-mission (two days), In-mission (6 days) and Post-mission (1 day). The plan represented a repeated-measures design (ABA) with two outside events (beginning and end of the isolation period). Since the control group (ground crew) did not experience the isolation phase, a between-group comparison would be possible to assess the effects of these outside events.

Within each session, control load was manipulated at two levels: *high* and *low* (4 vs 2 automatic controllers failing). High control load occurred either in the second or in the third trial, which was randomly determined by the simulation software. The first trial of each session always contained the low control load condition, which was treated as a warm-up for the operator before completing a trial with 'high' and ‘low’ control loads. The experimental design is shown in Figure 7.1.

Testing time	t1	t2	t3	t4	t5	t6	t7	t8	t9
Mission phase	Pre-M	Pre-M	In-M	In-M	In-M	In-M	In-M	In-M	Post-M
Exp. Group (cabin crew)	A	A	B	B	B	B	B	B	A
Control Group (ground crew)	A	A	A	A	A	A	A	A	A

Figure 7. 1: Experimental design (Pre-M = pre- mission phase, In-M = in- mission phase, Post-M = post- mission phase; B = exposure to isolation and confinement, A = non-exposure; within each experimental session: High and low control load)

It was planned to use a three-way analysis of variance with one between-subject variable (cabin vs ground crew) and two within-subject variables (control load, mission day). Control load was analysed at three levels and mission day at nine levels. Some of the data were transformed by means of either a logarithmic or a square root transformation to reduce variance heterogeneity. However, for the sake of clarity all means reported are untransformed values.

7.2.1.2 Participants

The participants in the study were four fully trained astronauts (three men and one woman) from the Canadian space programme and four members of ground control as a control group. Their ages ranged from 27 to 35. The members of the cabin crew are referred to as A, B, C and D.

7.2.1.3 Training and experimental task

In the first training session the basic principles of AMT were explained to the participants. After this session, training continued with self-administered sessions under the supervision of a designated crew member. The training plan envisaged extensive practice with a total of 5 training sessions. However, because of time constraints in the preparation programme, the number of training sessions actually completed varied between 2 and 4 sessions per crew member.

The experiment required a modification of the original AMT task since training time was limited, and access of the experimenter to the participants was restricted. Therefore, a version was required with a high level of reliability and a reduced level of complexity, which would not require the permanent presence of the experimenter during training. This resulted in a new version of the task (AMT 2.0, see Chapter 4), which emphasised fault detection and manual control skills.

7.2.2 Results

As pointed out in the description of the experimental design, it was planned that the ground crew also completed the task according to the schedule of the cabin crew. However, the data received from the ground crew were incomplete with only a few sessions fully accomplished. The ground crew claimed that they had been far too busy with their main tasks, hence, being unable to complete the AMT sessions as planned. The availability of only 5 complete experimental sessions (out of 36 scheduled ones) from ground personnel made it impossible to make any sensible comparison between ground and cabin crew. Therefore, the data collected from ground personnel were not included in this analysis. This unforeseen event weakened the design of this experiment since the existence of a control group (not exposed to isolation and confinement) would have provided a firmer basis for the evaluation of treatment effects.

However, even after the loss of the control group, the design did not differ from those of the two other field experiments, where the use of a control group was not possible. Therefore, two methods were employed to analyse the data. First, a regression line was fitted to aid visual inspection of the data in determining an approximate trend over the mission. Second, analyses of variance were carried out to look at effects of control load and mission phase across all crew members. Apart from the loss of data for the control group, the data set of the cabin crew was virtually complete, though in a small number of sessions the last one or two minutes of the third trial were missing.

A two-way within-subject ANOVA revealed a significant effect of control load and mission duration on most primary and secondary performance indicators. The effect of control load was found for all crew members. The results are reported below for each outcome measure.

7.2.2.1 Primary task performance

Parameter control failures

The overall effect of control load on the system control performance was very strong with performance suffering under high control (F=26.8; df=2,6; p<.005). Surprisingly, PCFs (parameter control failures) decreased with increased mission length, from 4.0% during pre-mission testing to 0.9% in the post-mission session (see Figure 7.2). This was found to be significant (F=7.5; df=8,24; p<.001). The decrease in errors was due to strong performance improvements for the HiCL (high control load) condition since only small improvements were observed for the two LoCL (low control load) conditions (see Figure 7.2). This interaction was highly significant (F=5.0; df=16,48; p<.001).

Fault diagnosis

Diagnostic activities as a function of control load showed that more incorrect diagnoses were made under HiCL (see Figure 7.3) but the effect proved not to be significant (F<1). Most errors tended to be committed in the pre-mission phase and at the end of the in-mission phase but no significant effect of mission duration was found (F=2.10; df=8,24; p>.05). No interaction was found. The number of incorrect fault diagnoses totalled only 15, virtually equally balanced between omissions and false alarms. In eight cases crew members did not recognise the failure of an automatic controller (omission) and in seven cases they wrongly reported a failed controller (false alarm).

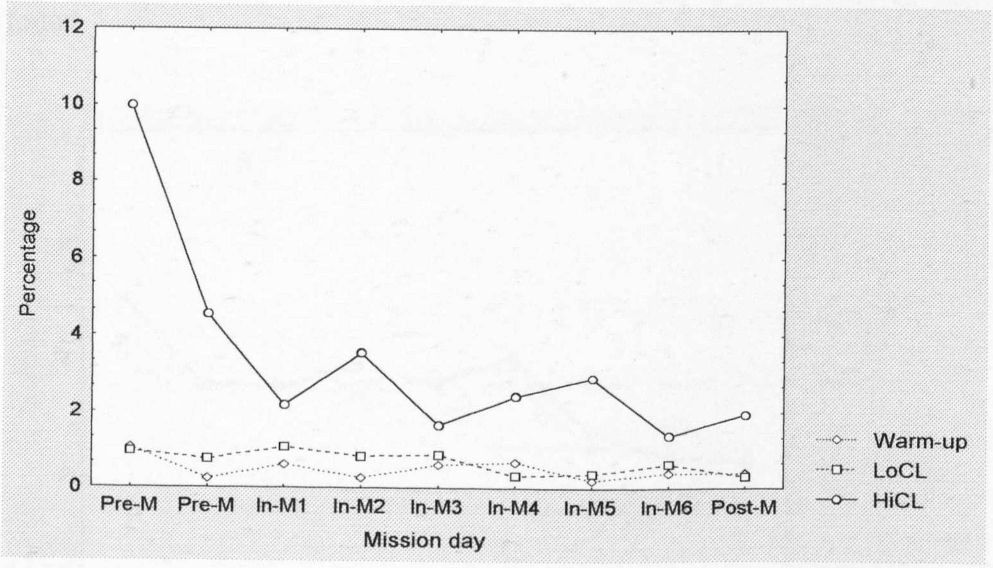


Figure 7. 2: Parameter control failures (%) as a function of control load and mission duration

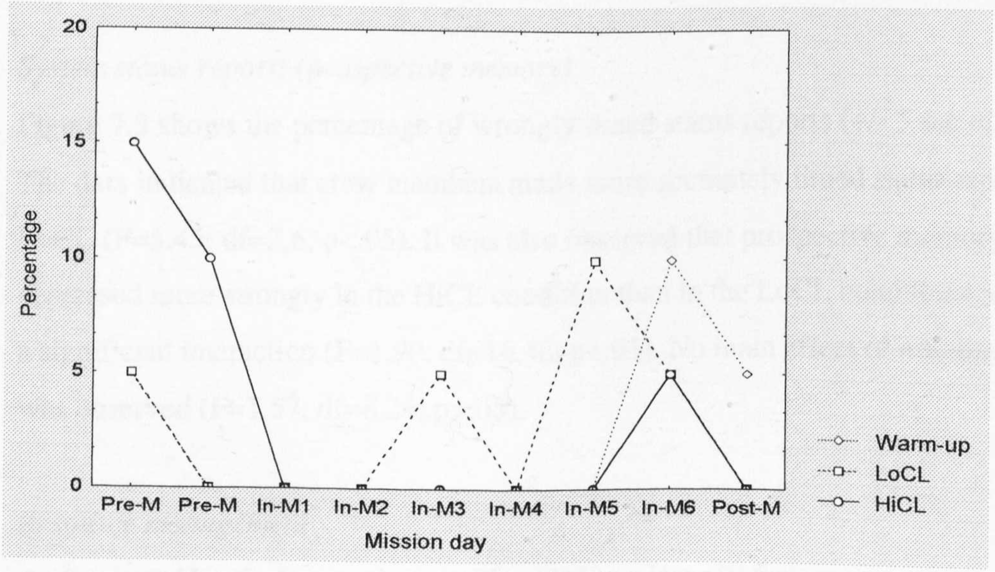


Figure 7. 3: Diagnostic errors (%) as a function of control load and mission duration

7.2.2.2 Secondary task performance

Reaction time

The results of the reaction time measure are presented in Figure 7.4. Although reaction time was somewhat higher under high control load, the effect did not reach statistical significance (F<1). However, a reduction in RT over the mission was recorded, with RT in the pre-mission phase being more than double (4.5 sec) than in the post-mission phase (2.2 sec). This effect was significant (F=5.8; df=8,24; p<.001). No interaction effect was found.

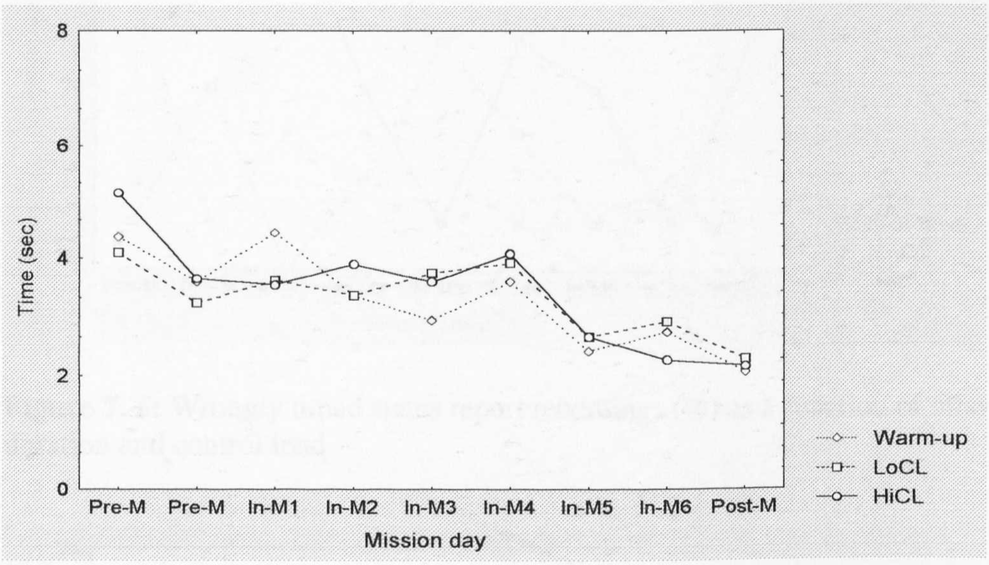


Figure 7. 4: Alarm reaction time (sec) as a function of control load and mission duration

System status reports (prospective memory)

Figure 7.5 shows the percentage of wrongly timed status reports (+/- 5 sec of time due). The data indicated that crew members made more accurately timed status reports during LoCL (F=5.45; df=2,6; p<.05). It was also observed that prospective memory errors decreased more strongly in the HiCL condition than in the LoCL conditions, resulting in a significant interaction (F=1.90; df=16,48; p<.05). No main effect of mission duration was observed (F=1.57; df=8,24; p>.05).

Resource management

An increased level of manual control load led to a less efficient management of resources, indicating that the control strategies employed by the crew members became less effective (see Figure 7.6). This observation was highly significant (F=61.1; df=2,6; p<.001). Performance also improved on this measure with increasing mission length (F=4.2; df=8,24; p<.005). The improvement in performance was more pronounced in the HiCL condition, resulting in a significant interaction (F=2.1; df=16,48; p<.05).

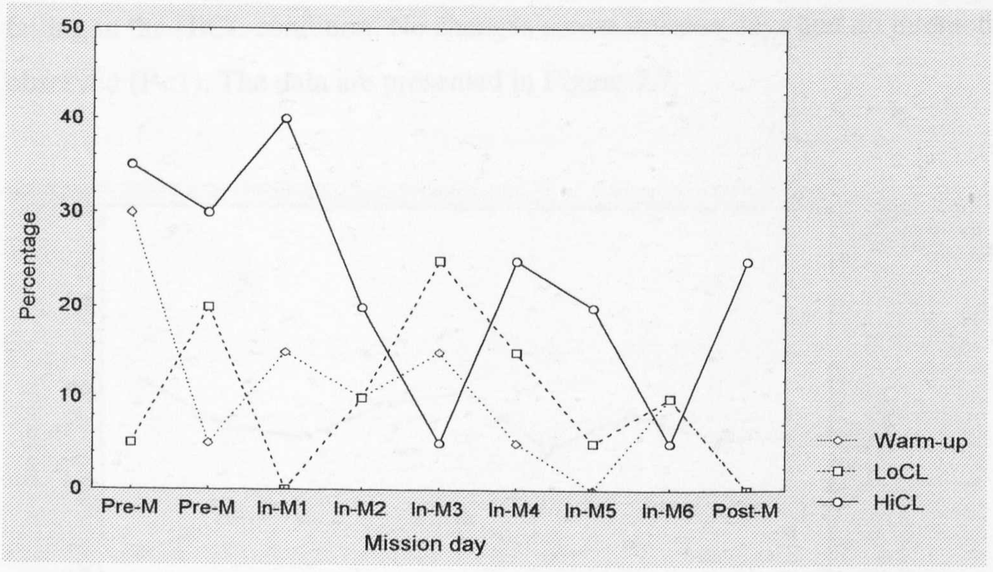


Figure 7. 5: Wrongly timed status report recordings (%) as a function of mission duration and control load

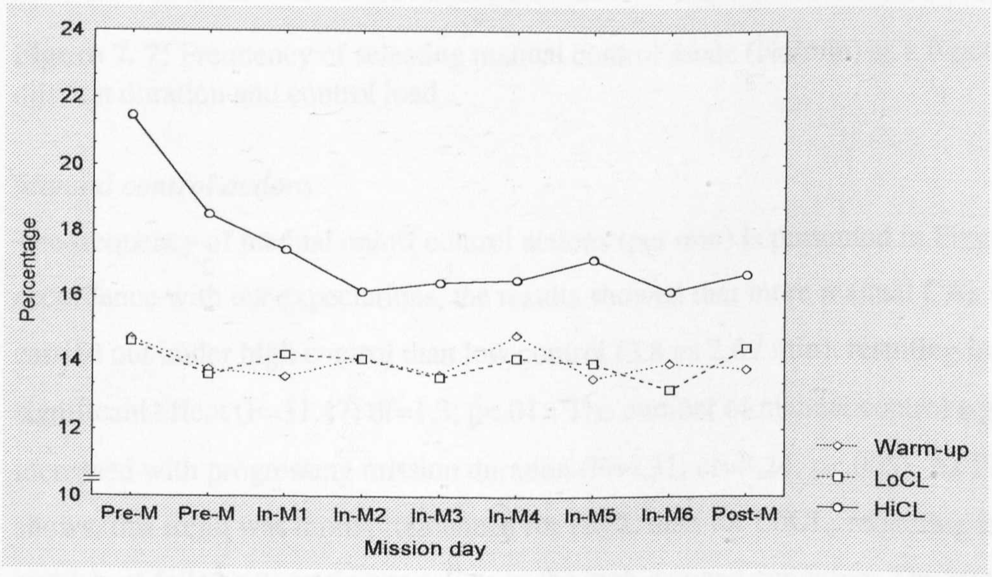


Figure 7. 6: Resource consumption (%) as a function of mission duration and control load

7.2.2.3 Manual control activities

Use of manual control mode

An analysis of manual control activities was only carried out for the second LoCL (the first LoCL scenario was not analysed) and the HiCL conditions. As expected, it showed that crew members changed more frequently from AUTO to MAN control under HiCL (0.54/min) than LoCL (0.29/min), resulting in a significant effect (F=122.1; df=1,3;

p<.005). This was largely a reflection of twice the number of automatic controllers failing in the HiCL condition. No changes across mission days and no interaction were observed (F<1). The data are presented in Figure 7.7.

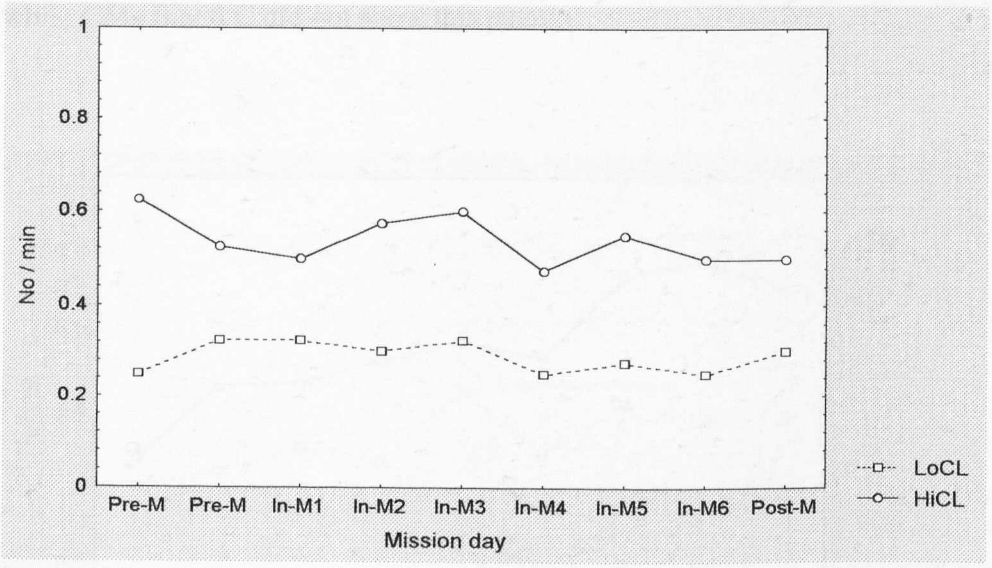


Figure 7. 7: Frequency of selecting manual control mode (No/min) as a function of mission duration and control load

Manual control actions

The frequency of manual on/off control actions (per min) is presented in Figure 7.8. In accordance with our expectations, the results showed that more manual CAs were carried out under high control than low control (3.8 vs 2.6 / min), resulting in a significant effect (F=51.47; df=1,3; p<.01). The number of manual control actions also increased with progressing mission duration (F=4.31; df=8,24; p<.005). As Figure 7.8 shows, this trend was more pronounced for HiCL than for LoCL, reflecting the larger number of failed automatic controllers in the high demand condition. The interaction was significant (F=2.57; df=8,24; p<.05).

7.2.2.4 Subjective state

The rise in task demands during the HiCL condition resulted in an increase in task-related effort to maintain performance levels (see Figure 7.9). This increase was significant (F=7.27; df=2,6; p<.05). As the mission progressed, the gap between effort expended on the HiCL condition and the two LoCL conditions became smaller, resulting in a significant interaction effect (F=1.97; df=16,48; p<.05). Figures 6.10 and 6.11

present the data for anxiety and fatigue as a function of control load and mission duration. The data analysis did not reveal any significant effect on anxiety or fatigue (all F<1). For two crew members (A and D) a considerable increase in effort expenditure was observed during in-mission testing compared to pre- and post-mission baselines while CMs B and C did not show this pattern.

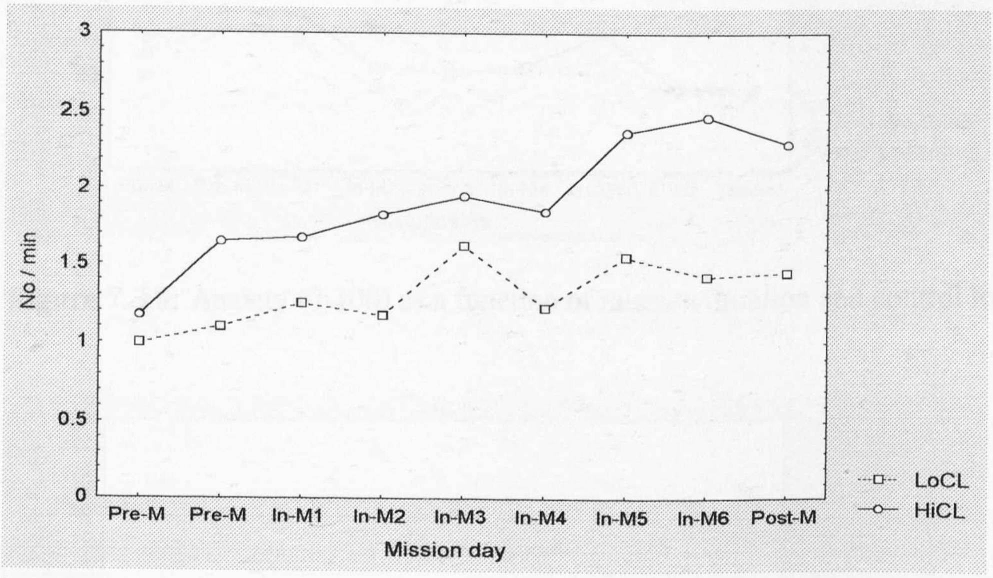


Figure 7. 8: Frequency of manual on/off control actions (No/min) as a function of mission duration and control load

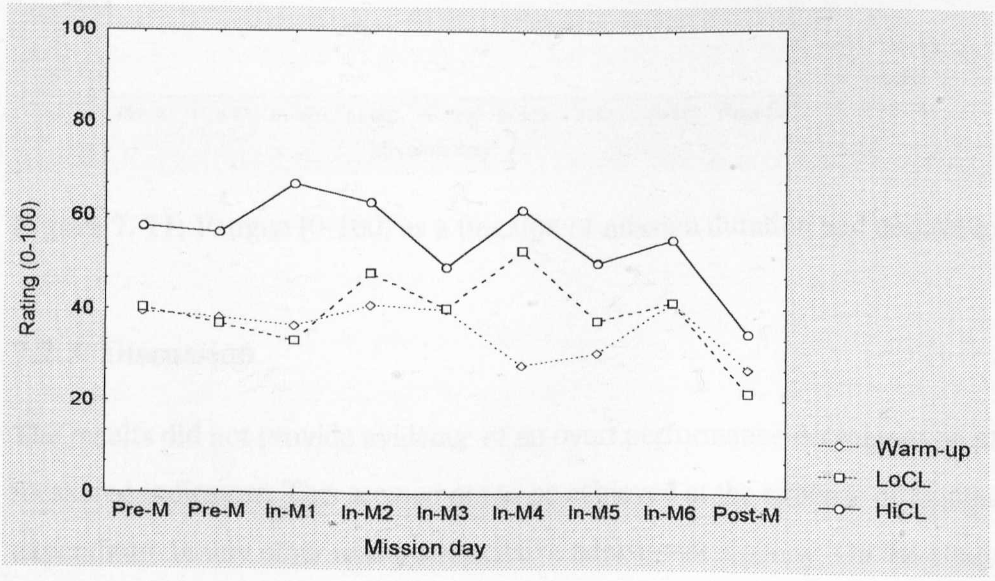


Figure 7. 9: Effort (0-100) as a function of mission duration and control load

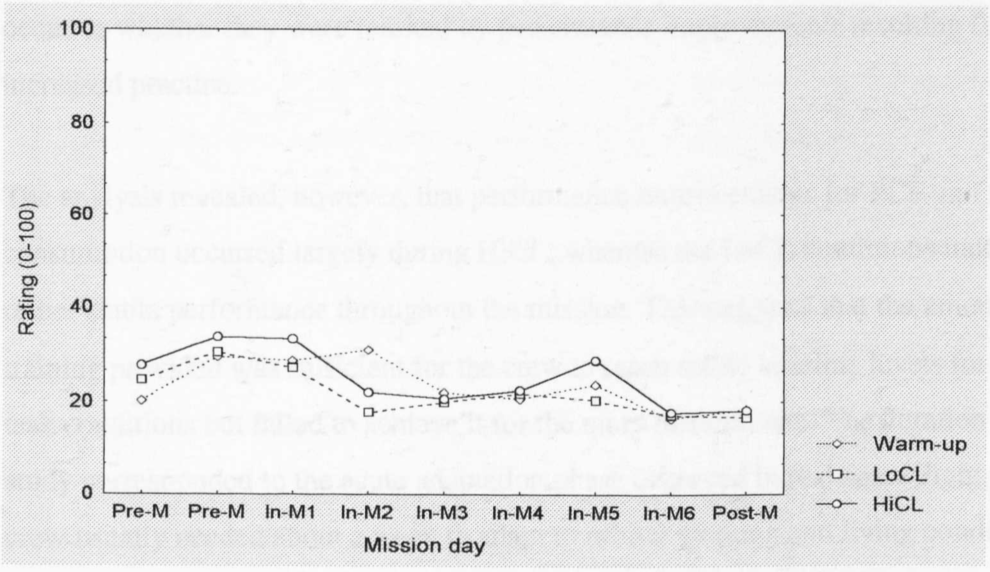


Figure 7. 10: Anxiety (0-100) as a function of mission duration and control load

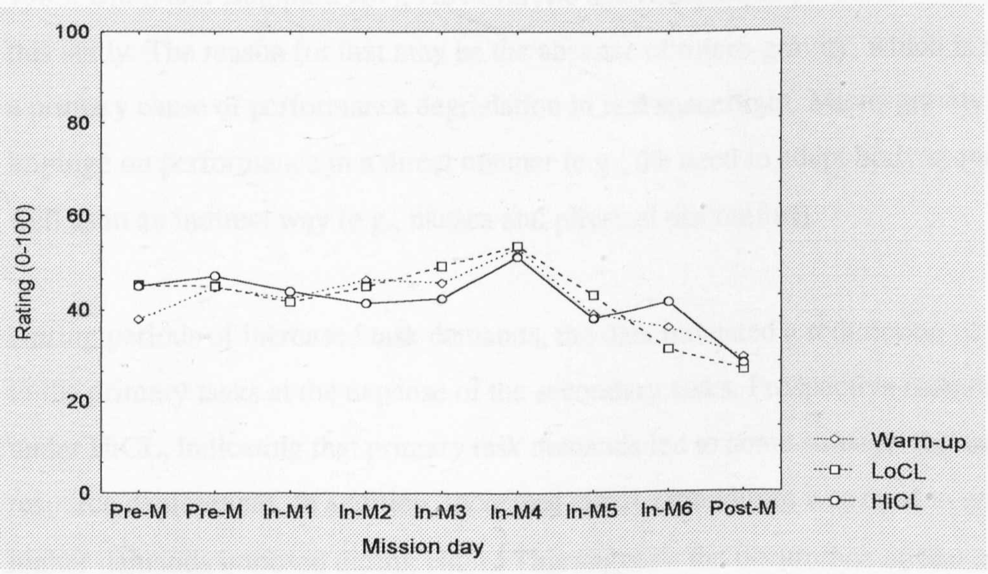


Figure 7. 11: Fatigue (0-100) as a function of mission duration and control load

7.2.3 Discussion

The results did not provide evidence of an overt performance decrement on any of the measured indicators. This seemed not to be achieved at the expense of increased effort expenditure or any other resource-intensive adjustment strategy. On the contrary, crew members tended to improve their performance with increasing mission length. This was probably due to a practice effect since the amount of training received was very unlikely to have been sufficient to achieve stable performance levels. However, in the absence of a control group, it is difficult to say whether performance decrements did really not occur or whether they were masked by performance improvements resulting from increased practice.

The analysis revealed, however, that performance improvements for PCF and resource consumption occurred largely during HiCL, whereas the LoCL conditions indicated a rather stable performance throughout the mission. This suggests that the amount of training provided was sufficient for the crew to reach stable baseline levels for the easier task conditions but failed to achieve it for the more difficult one. The duration of this study corresponded to the acute adaptation phase observed in real spaceflight, where the crew usually needed about a week to adapt to orbital working and living conditions (Gushin et al 1993). The first few days of a space mission have commonly been found to show a decline in performance compared to pre-mission baseline levels (Connors et al 1985; Bluth and Helppie 1987). However, no such adaptation problems were evident in this study. The reason for that may be the absence of micro-gravity, which is regarded as a primary cause of performance degradation in real spaceflight. Micro-gravity may impinge on performance in a direct manner (e.g., the need to adapt body movements) as well as in an indirect way (e.g., nausea and physical discomfort).

During periods of increased task demands, the data indicated a redirection of resources to the primary tasks at the expense of the secondary tasks. Prospective memory suffered under HiCL, indicating that primary task demands led to some strategic adjustments in resource deployment. In addition, increased effort expenditure was used to cope with the higher demands imposed during HiCL. This suggests the occurrence of two of the strategies (strategic adjustment and compensatory costs) outlined in the theoretical framework. As the mission progressed, the gap between HiCL and LoCL disappeared completely for prospective memory (on testing day 4) and it became smaller for task-related effort. This indicates that primary task requirements associated with HiCL diminished considerably with increasing mission length but did not vanish entirely, even after extensive practice. The observed improvements in control performance over the mission were accompanied by an increase in manual control activities. Although it is not possible to determine whether a causal relationship between the two observations exists, some indirect evidence was found. At the beginning of the mission little difference was observed between control load conditions, though four failed automatic controllers would be expected to require approximately twice as many manual CAs than two failures. At the end of the mission, the observed ratio of manual CAs between HiCL and LoCL was very close to the expected 2:1 ratio. Therefore, it seems reasonable to assume that increased manual control activity resulted in improved performance (though, it is difficult to determine the optimal number of manual CAs under HiCL). This suggests that a lack of manual control activity was responsible for poorer control performance in the early mission phases rather than the use of ineffective control actions that failed to bring the system back into the target state. This is also supported by the frequency of use of the manual control mode, where little change was recorded over the course of the mission. The crew entered the manual control mode but failed to carry out appropriate on/off CAs to manage the system effectively.

The finding that stable baseline levels are reached more quickly with simpler task levels (i.e. low control load in the present case) also has some implications for mission simulations of this kind. If the mission simulation schedule does not allow for extensive practice before mission start, reducing the task difficulty within the same task environment may be an alternative. This is because the attainment of stable baselines permits easier detection of any decrements as a function of ICE or other forms of sub-optimal working conditions than if this judgement has to be made while performance is still subject to continuous learning. However, the disadvantage of reducing task difficulty is that it entails a reduction in sensitivity of the task environment to operator strain and may therefore fail to detect more subtle forms of performance decrement.

The results also indicated that the performance measures showed a more consistent trend across crew members than the subjective state measures, where interindividual differences were much more pronounced. These interindividual differences in subjective state suggested that some strain was observed for two crew members in the form of increased effort and anxiety levels during the in-mission phase. However, this did not become evident in the overall group mean. Interindividual differences are an important issue in space research but a detailed analysis of these was not the primary goal of this thesis since it has adopted a normative approach.

7.3 Experiment 5

The experiment was carried out during an over-wintering expedition in the Antarctic, which was partly funded by ESA. A group of French scientists and technical support personnel spent the Antarctic winter in their camp in Adélie Coast while pursuing scientific investigations. After having been instructed and trained on the AMT simulation task, a co-investigator supervised the completion of the experimental session in the Antarctic.

During the expedition the participants were repeatedly tested on AMT and a number of other performance tests (not a subject of this thesis) to address the effects of long-term isolation and confinement. Completed over a time period of 8 months, it was the experiment with the longest period of exposure to ICE in the thesis. Overall, conditions of isolation and confinement were somewhat less severe than in a typical space station. The habitat was of a larger size and there were more opportunities to leave the camp for a short period of time, though not during the Antarctic winter. Social isolation was also less severe since a total of 32 expedition members stayed in the camp (of which 16 took part in the study).

As in the previous experiment, it was hypothesised that under high control load secondary task performance would deteriorate and effort expenditure would be increased to cope with the augmented task demands. Furthermore, it was expected that performance and possibly subjective state measures would become more unstable in the second mission half as a result of the adaptation processes described by the in-mission phase model. The Antarctic winter is also considered an important factor since it does not allow the expedition members to leave the camp.

7.3.1 Method

7.3.1.1 Design and treatment of data

The study used repeated-measures design (AB-design) with control load as an additional within-subject variable (see Figure 7.12). It was planned to test the participants once before and 8 times during the isolation period. The set-up of the experimental sessions was identical to the one in Experiment 4, with the expedition crew completing a 30-min session, consisting of three 10-min trials. The data were analysed in the same way as in Experiment 4 (see section 7.2.2.1).

Testing time	t1	t2	t3	t4	t5	t6	t7	t8	t9
Mission phase	Pre-M	In-M	In-M	In-M	In-M	In-M	In-M	In-M	In-M
Exp. Group	A	B	B	B	B	B	B	B	B

Figure 7. 12: Experimental design (Pre-M = pre-mission phase, In-M = in-mission phase; B = exposure to isolation and confinement, A = non-exposure; within each experimental session: High and low control load)

7.3.1.2 Participants

A total number of 16 male participants took part in this experiment. The were all members of a French over-wintering expedition. The study participants can be divided into two groups according to their professional background: scientific (10) and technical personnel (6). Their ages ranged from 23 to 51 years although most participants were under 30 years of age (median =25).

7.3.1.3 Training and experimental task

As in experiment 4, AMT 2.0 was used as the experimental task environment. Participants completed between 2 and 3 training sessions before embarking on their voyage to the Antarctic. Training was administered by a co-instructor who was sufficiently familiar with the software. Compared to the other versions of the simulation task, AMT 2.0 requires less training time to acquire a sufficient level of competence.

7.3.2 Results

A considerable number of missing data (only 10 expedition members completed their testing sessions) from the pre-mission phase did not allow the inclusion of the pre-mission testing session in the analysis. A descriptive analysis of the data revealed remarkable differences between the professional groups (scientists and technicians) on all performance measures. This made it impossible to treat the 16 participants as a homogeneous group and required the introduction of an additional between-subject variable: professional group.

7.3.2.1 Primary task performance

Parameter control failures

The data of system control performance is summarised in Figure 7.13. It indicates a considerable difference in performance between the two groups with the scientists (SCI) committing significantly fewer errors (F=21.6; df=1,11; p<.001). An analysis of the changes over the expedition period did not reveal any significant trend (F<1). However, control load showed a very strong effect with the number of control errors increasing dramatically under HiCL (F=59.1; df=2,22; p<.001). The effect of poorer control performance under HiCL was more pronounced for the technicians (TEC) than for the SCI-group, resulting in a significant interaction (F=9.19; df=2,22; p<.005). No other significant effect was found.

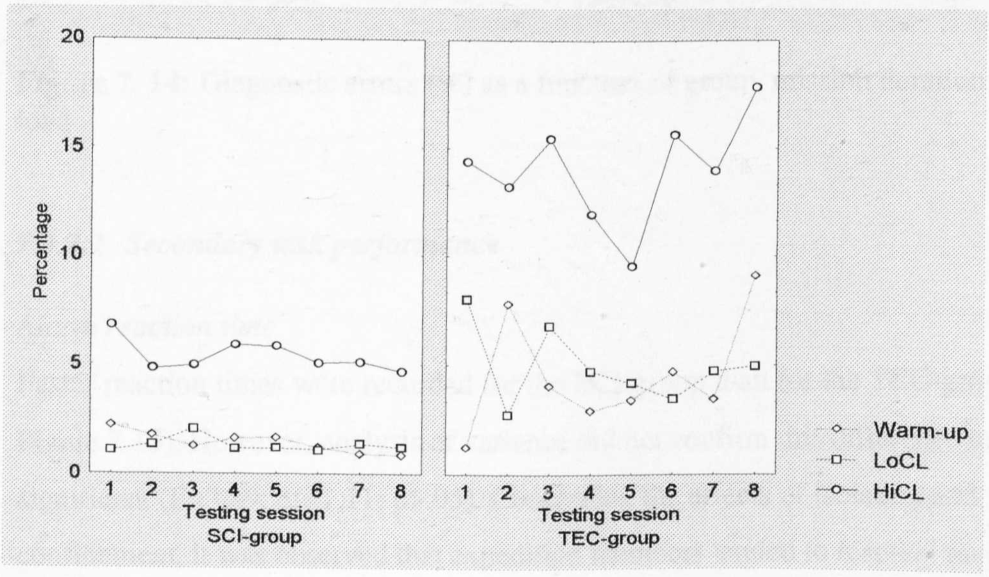


Figure 7. 13: Parameter control failures (%) as a function of group, mission duration and control load

Fault diagnosis

The data for this measure are shown in Figure 7.14. As observed for PCF, this primary task measure also showed a better performance for the scientists (error rate: 4.4 %) than for the technical support personnel (28.4 %). The difference was statistically significant (F=8.13; df=1,11; p<.05). Furthermore, a linear trend of performance improvement with increasing mission length was observed for both groups (F=2.17; df=7,77; p<.05). During HiCL the number of misdiagnoses increased significantly compared to the low control condition. This difference was found to be significant (F=4.32; df=2,22; p<.05). No interactions were observed.

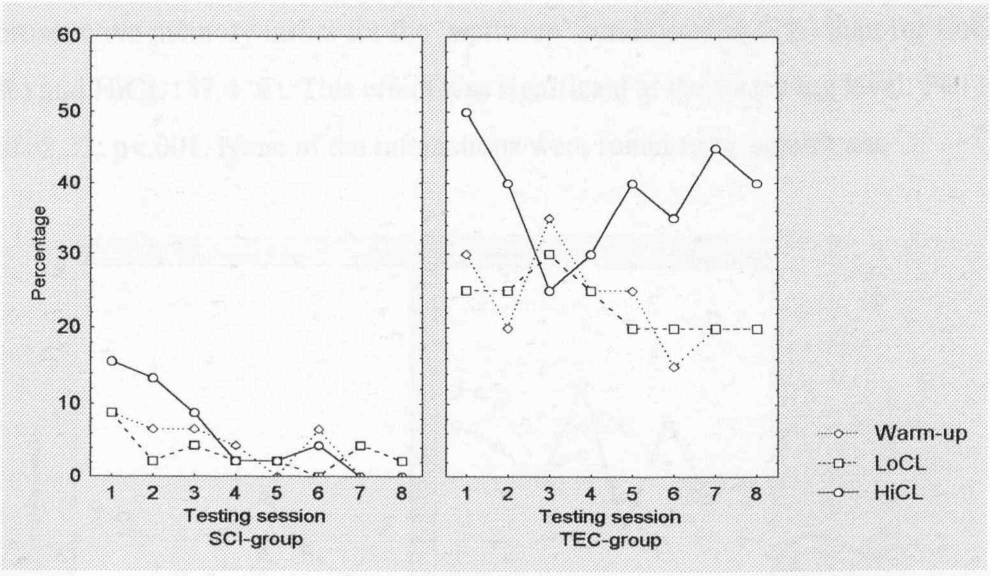


Figure 7. 14: Diagnostic errors (%) as a function of group, mission duration and control load

7.3.2.2 *Secondary task performance*

Alarm reaction time

Faster reaction times were recorded for the SCI-group than for the TEC-group (see Figure 7.15). However, analysis of variance did not confirm this difference to be significant (F=3.89; df=1,11; p>.05). Concerning the effects of isolation and confinement, it was observed that expedition members tended to respond more quickly to the warning signals as the over-wintering expedition progressed. This improvement of reaction time performance was significant (F=2.56; df=7,77; p<.05). Although reaction times were longer, as predicted, for HiCL, the effect failed to reach significance levels (F=2.95; df=2,22; p>.05). This was because control load only affected the scientists in the predicted direction whereas no effect was observed for the TEC-group. The interaction was significant (F=3.53; df=2,22; p<.05). No further interaction was found.

System status reports (prospective memory)

The data were scored for timing errors, i.e. recordings made more than +/- 5 sec outside the required times were counted as wrongly timed. Again, the scientists showed better performance on this task than the TEC-group (see Figure 7.16 for the data). However, the difference did not reach statistical significance (F=1.13; df=1,11; p>.05). The analysis of mission phases did not reveal any clear trend (F=1.21; df=7,77; p>.05). The only significant main effect observed referred to control load, which showed fewer prospective memory errors for the 'warm-up' condition (28.0 %) than for LoCL (35.9 %) and HiCL (37.4 %). This effect was significant at the following level: F=11.51; df=2,22; p<.001. None of the interactions were found to be significant.

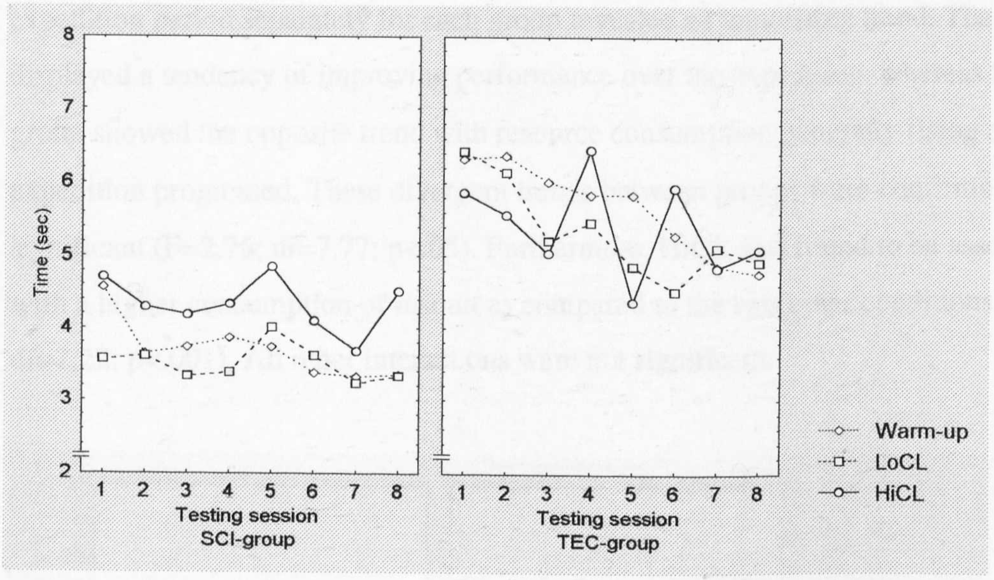


Figure 7. 15: Reaction time (sec) as a function of group, mission duration and control load

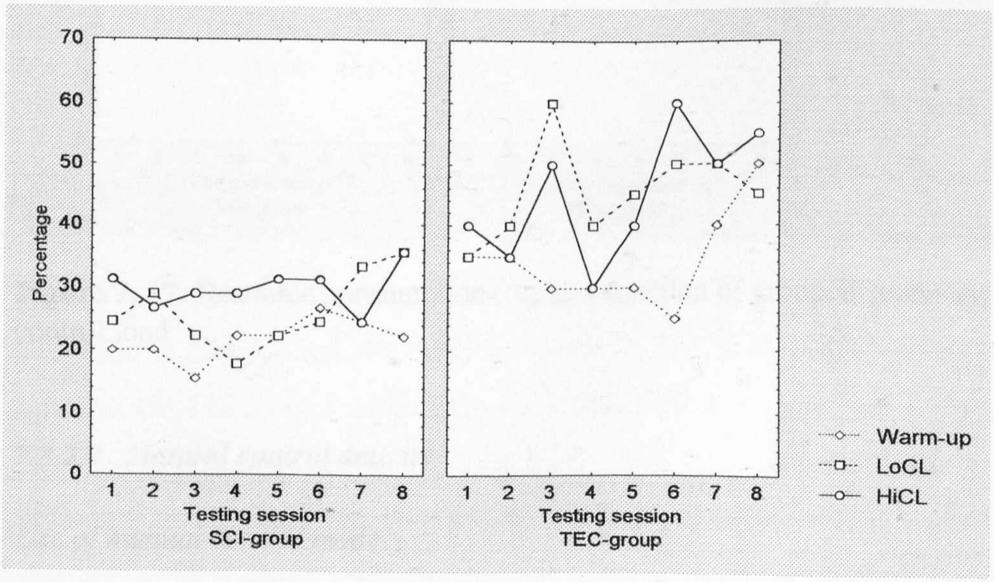


Figure 7. 16: Wrongly timed SSR (%) as a function of group, mission duration and control load

Resource management

This measure indicates the efficiency of the system control performance rather than its effectiveness (which is measured by PCF). The results (see Figure 7.17) suggested that more economical control strategies were employed by the scientists compared to the support personnel. However, this difference failed to reach significance levels (F=4.1; df=1,11; p>.05). With regard to the effect of ICE, no clear trend was observed for the overall group (F=2.01; df=7,77; p>.05). However, examining the changes over the expedition period separately for each group revealed an interesting trend. The SCI-group displayed a tendency of improving performance over the expedition, whereas the TEC-group showed the opposite trend with resource consumption generally rising as the expedition progressed. These divergent trends between groups were confirmed to be significant (F=2.76; df=7,77; p<.05). Furthermore, HiCL was found to be associated with a higher consumption of resources compared to the two other conditions (F=70.1; df=2,22; p<.001). All other interactions were not significant.

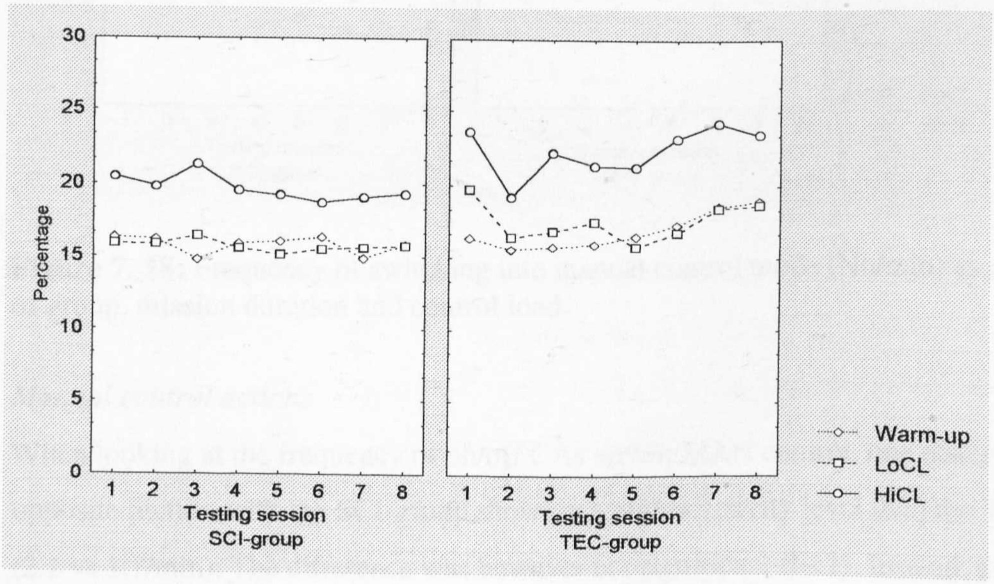


Figure 7. 17: Resource consumption (%) as a function of group, mission duration and control load

7.3.2.3 *Manual control activity*

Use of manual control mode

An analysis of manual control activities was only carried out for the second LoCL trial (i.e. the first LoCL scenario was not analysed) and HiCL. It showed that the SCI-group switched less frequently between manual and auto control (0.58 vs 0.78/min) than the

TEC-group. However, the effect was not significant (F=1.42; df=1,11; p>.05). Over the mission a linear reduction in the number of mode changes was observed for both groups (see Figure 7.18), indicating a more optimal use of the manual control mode. The trend showed overall significance (F=7.46; df=7,77; p<.001) and planned comparisons confirmed the presence of a linear trend (F=9.51; df=1,11; p<.01). As expected, a larger number of mode changes was observed under HiCL (0.78/min) than LoCL (0.58/min). This difference was highly significant (F=39.9; df=1,11; p<.001).

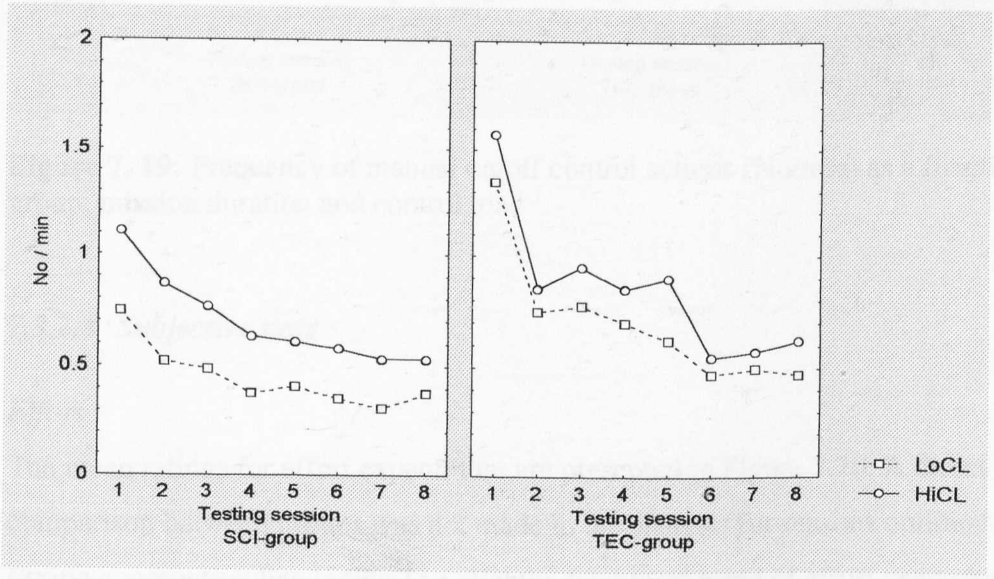


Figure 7. 18: Frequency of switching into manual control mode (No/min) as a function of group, mission duration and control load

Manual control actions

When looking at the frequency of on/off CAs within MAN control, one observed the opposite picture with the SCI-group showing a higher activity level than the TEC-group (2.1 vs 1.7/min). The difference was however not significant (F<1). Instead, a significant interaction between group and control load was found (F=8.78; df=1,11; p<.05). This was because of a stronger rise in CAs under HiCL (compared to LoCL) for the scientists (+ 55%) than for the technicians (+21%). A main effect of control load was also found in the predicted direction (HiCL: 2.1/min; LoCL: 1.6/min). It was significant at the following level: F=36.6; df=1,11; p<.001. No effect of mission duration was found (F=1.50; df=7,77; p>.05). The data for on/off MCA are presented in Figure 7.19.

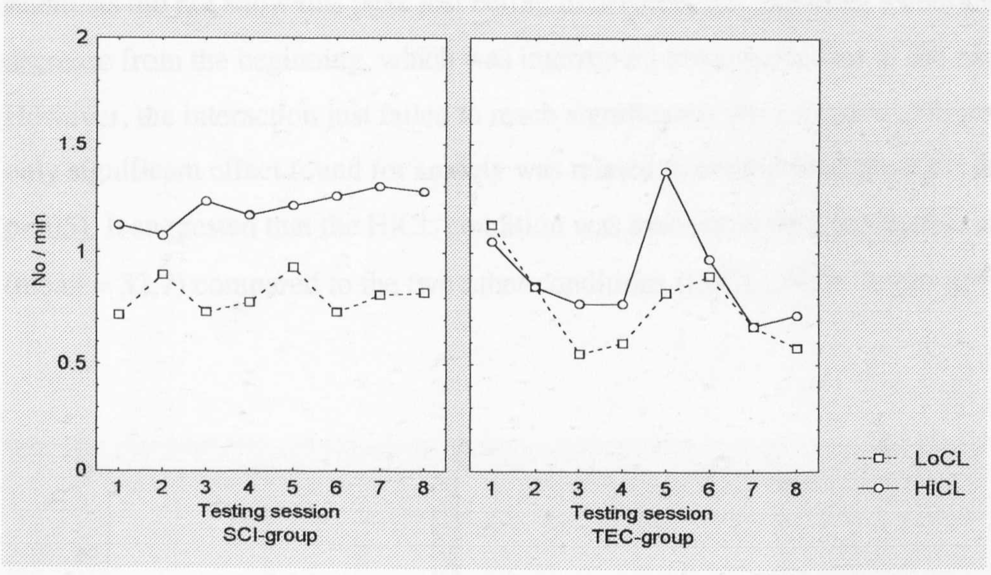


Figure 7. 19: Frequency of manual on/off control actions (No/min) as a function of group, mission duration and control load

7.3.2.4 Subjective state

Effort

The mean ratings for effort expenditure are presented in Figure 7.20. A direct comparison between groups was not made in the analysis for reasons outlined above. Fitting a regression line indicated a slightly downward trend of effort expenditure for both groups, with a plunge after the 5th testing session, possibly associated with the end of the Antarctic winter. This reduction was more pronounced for the TEC-group than for the scientists. Analysis of variance confirmed a significant effect at the following level: F=2.41; df=7,70; p<.05. Control load also showed an effect on effort with levels being highest for HiCL (mean: 42.3), followed by (mean: 38.8) and 'warm-up' (mean: 34.8). Each mean was significantly different from the others (LSD-test: min. p<.05). The significance level of the overall effect was: F=3.96; df=2,20; p<.05. No other interactions were found.

Anxiety

As the mean anxiety ratings in Figure 7.21 indicate, no clear trend in the rating pattern was observed over the expedition period (F=1.61; df=7,70; p>.05). A slightly different pattern between scientists and technicians was observed, with the TEC-group showing a continuous reduction in ratings after their peak three months into the mission. The scientists did not show this peak just before mid-winter but exhibited a continuous decrease from the beginning, which was interrupted towards the end of the expedition. However, the interaction just failed to reach significance (F=1.91; df=7,70; p>.05). The only significant effect found for anxiety was related to control load (F=4.47; df=2,20; p<.05). It suggested that the HiCL condition was associated with increased anxiety (mean = 32.7) compared to the two other conditions (LoCL: 28.5, 'warm-up': 28.0).

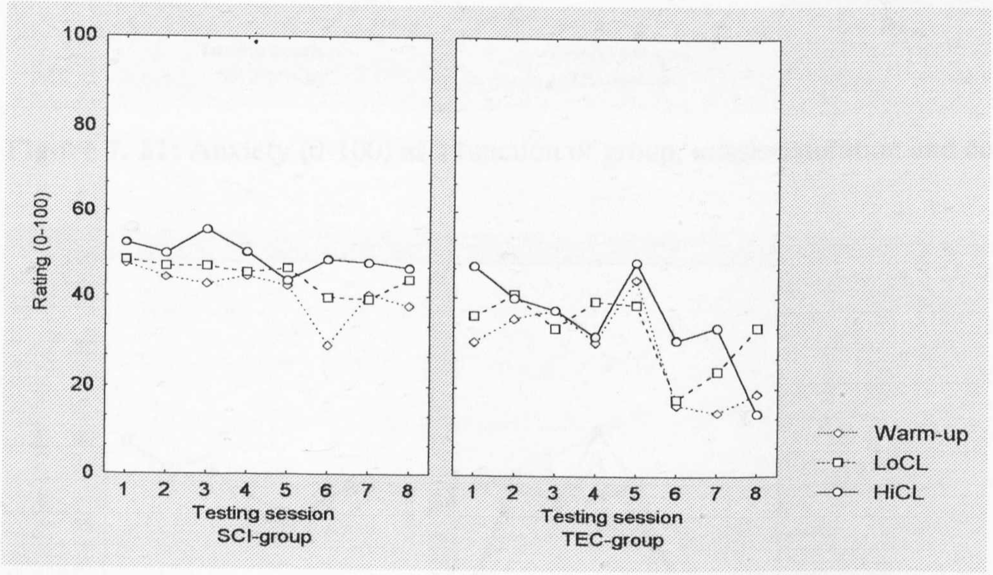


Figure 7. 20: Effort expenditure (0-100) as a function of group, mission duration and control load

Fatigue

As can be seen in Figure 7.22, no common pattern of fatigue ratings across groups was recorded (F=2.07; df=7,70; p>.05). However, each professional group showed a distinctly different rating pattern, which was reflected in a significant interaction between group and expedition phase (F=2.29; df=7,70; p<.05). The technical support personnel showed an inverted U-shaped curve with fatigue peaking at the height of the Antarctic winter. This was in contrast to the scientists, whose pattern of results may be described as a linear decrease, of which the trend was broken after the end of the mid-winter phase. The analysis did not reveal any effect of control load or any other interactions.

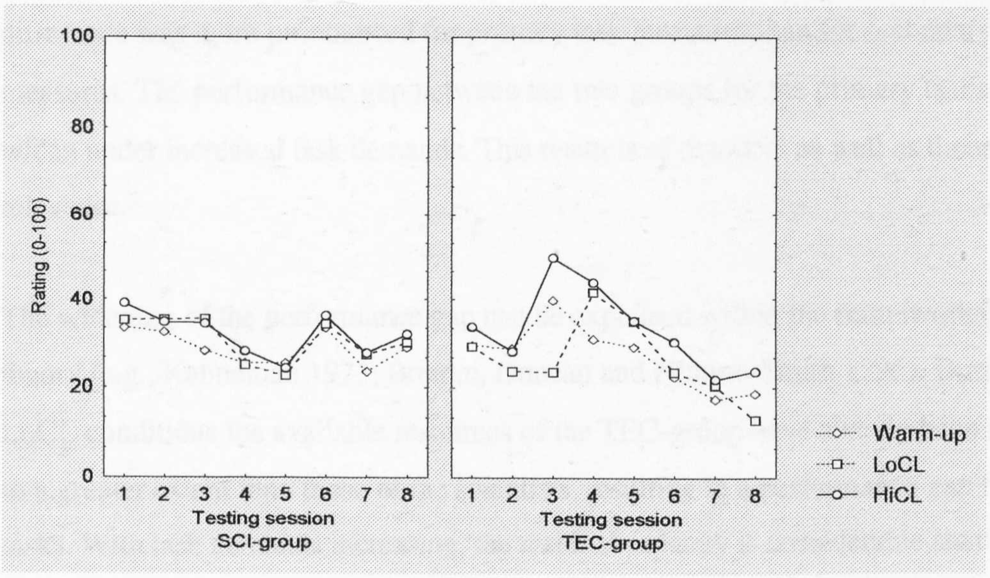


Figure 7. 21: Anxiety (0-100) as a function of group, mission duration and control load

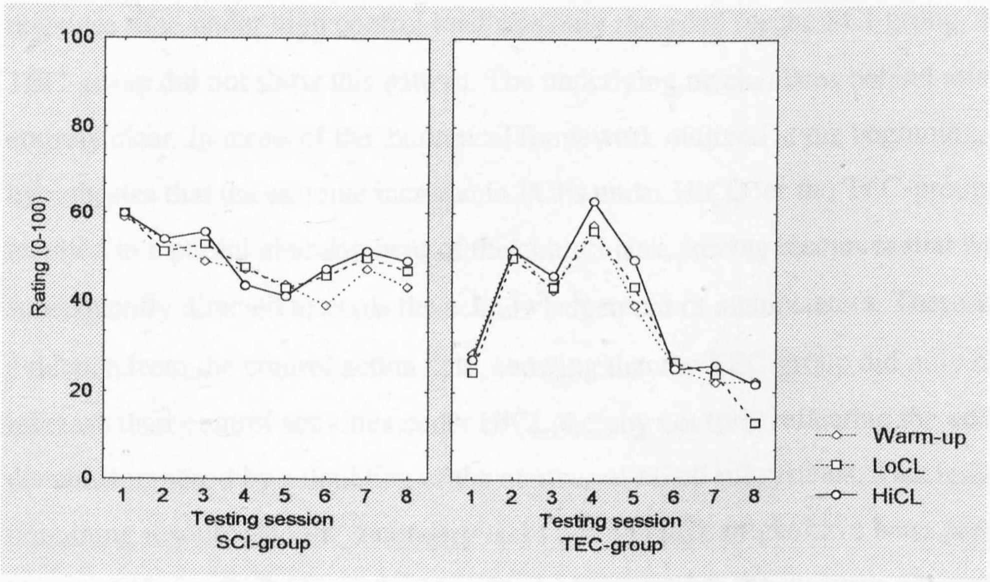


Figure 7. 22: Fatigue (0-100) as a function of group, mission duration and control load

7.3.3 Discussion

Overall, the results suggested an absence of any general decrement in performance during Antarctic over-wintering. However, changes in subjective state measures were observed over the course of the isolation period. These patterns are now discussed in more detail.

Remarkable differences in performance were found between the two groups (scientific and technical support personnel) with the scientists performing better on all tasks. The difference was more pronounced for primary task measures than for secondary task measures. The performance gap between the two groups for the primary tasks tended to widen under increased task demands. This result is of practical as well as theoretical relevance.

The widening of the performance gap can be explained within the framework of resource theory (e.g., Kahneman 1973; Bourke, Duncan and Nimmo-Smith 1996). During the LoCL conditions the available resources of the TEC-group have already been stretched to a greater extent than those of the scientists, resulting in a performance gap for primary tasks. With task demands increasing, the residual capacity is considerable smaller for the technicians, resulting in a stronger effect of performance decline for this group than for the scientists. In this context it is noteworthy that for reaction time an increase in response time under high control load was only recorded for the SCI-group, whereas the TEC-group did not show this pattern. The underlying mechanisms behind this are not entirely clear. In terms of the theoretical framework outlined at the beginning, one may hypothesise that the extreme increase in PCFs under HiCL for the TEC-group may have resulted in a partial abandonment of the control task, freeing resources that were subsequently directed towards the acknowledgement of annunciators. There was some evidence from the control action data, showing that the TEC-group did only marginally increase their control activities under HiCL, thereby not fully reflecting the additional demands imposed by a doubling of the number of failed subsystems. This reallocation of remaining resources to the secondary tasks during HiCL might have been part of an attempt to improve the time-sharing of tasks by changing the resource allocation policy (Norman and Bobrow 1975). Their performance resource functions might have been such that available resources were insufficient to have a noticeable impact on primary task performance but sufficient to maintain secondary task performance. Although this was in direct violation of natural task priorities (and experimental instructions), this phenomenon of dealing with non-central aspects of the task while neglecting the main goal has been observed previously (Dörner 1987).

The overall better performance of the scientists is probably a reflection of their superior mental abilities. This is considered to be a reasonable assumption, though it cannot be independently verified. Furthermore, motivational factors may have played a role. The technicians (or at least some of them) may have lacked sufficient motivation to make full use of their abilities on the (imposed) task, which they may have perceived as little relevant to their normal work duties.

It is acknowledged that explanations of this kind are merely post-hoc and not verifiable. However, from a practical point of view, this has important implications. Under normal operational conditions the crew members of higher ability may only show slightly better performance than their less able colleagues but under extreme conditions (e.g. emergencies, increase in overall workload) this difference in performance may widen to a critical gap.

In addition to the interaction between group and control load, the study also showed a considerably main effect of control load on most performance indicators, notably PCF and resource consumption, where performance degradation under increased control load were observed for both groups. This was consistent with the findings of Experiment 4 where high control load represented such a considerable increase in task demands that it was not possible to maintain normal performance levels for parameter control. Interestingly, prospective memory showed no clear effect of control load in this study. This is inconsistent with the findings in Experiment 4 where prospective memory errors increased under HiCL. This seems to indicate that concentration of available resources on certain tasks is subject to considerable variation, even under similar environmental conditions, with crew members focusing on certain secondary tasks while neglecting others during periods of increased demands.

Again, increased effort expenditure was associated with high levels of control load although the effect was not as strong as in the previous experiment. Effort was also found to be significantly correlated with performance measures for a number of expedition members, pointing into the same direction like the findings from Experiment 4. It suggested (although a causal link is not certain) that crew members tended to put more effort into the task when their performance was at risk.

Examining the main effect of mission duration did not reveal the commonly found performance improvements, which is expected in situations where participants have had insufficient training to reach stable baseline levels of performance (as it was the case in this study). Only two measures showed some trend of improving performance with increasing expedition duration whereas the others showed remarkable stability during the study. There may be two explanations for the absence of general performance improvements. First, the strain experienced during the over-wintering expedition led to performance decrements that masked any practice effects. Second, because of the long intervals between testing sessions (up to 2 months), practice effects were rather small. In the absence of a control group, it is difficult to determine which explanation is more accurate. However, the second explanation is to be favoured, since many studies have indicated the stability of performance on the main task in the face of increased strain (e.g., Hockey 1993).

The trend of decreasing fatigue levels during the in-mission phase, which had already been observed in the previous study, occurred again for the SCI-group. In contrast, after an initial decline in anxiety and fatigue at the start of the in-mission phase, the TEC-group showed a mid-mission increase in both measures. This increase occurred during the height of the Antarctic winter. Confinement levels may have been aggravated during this time because of the impossibility of leaving the camp. This may have affected the TEC-group more strongly because of their different work activities. These observations are consistent with those from previous over-wintering studies, which have found effects of negative mood, sleep disturbances, and increased anxiety and fatigue during the Antarctic winter (Oliver 1990; Rivolier, Cazes and McCormick 1990). Similar patterns have also been observed in space-related isolation studies (Bluth and Helppie 1987; Gushin et al 1993; Connors et al 1985).

7.4 Experiment 6

This experiment was carried out during a simulation of a long-duration spaceflight (HUBES, see section 7.1.1) at the Institute for Biomedical Problems in Moscow. The spaceflight simulation, financed by ESA, took place in the Mir simulator. A crew of three was isolated for 135 days, during which a payload of 30 experiments was completed. The study allowed the crew to be tested at short and regular intervals over an extended period of time. Furthermore, the technical and ambient environment had a high degree of similarity with real spaceflight because the simulation took place in a Mir space capsule. The spaceflight simulation also served as preparation of the EuroMir-95 flight.

The conceptual distinction between practised and novel faults allowed the experimenter to look at the long-term development of implicit knowledge and performance for these fault types. It was expected that performance would be better during practised faults than novel faults and that this difference would gradually vanish as the mission progressed. It was also hypothesised that performance decrements, if observed, would be more pronounced for secondary tasks than for primary tasks.

7.4.1 Method

7.4.1.1 Design and treatment of data

The study design used multiple observations, allowing us to discover any changes in the outcome variables over the course of the mission. Measures were taken three times a week over the entire isolation period of 19 weeks. In addition, two pre- and two post-mission measurements were scheduled, resulting in a total number of 61 measurements (see Figure 7.23). Furthermore, within each testing session, crew members were confronted with three levels of fault difficulty to allow a comparison between different levels of cognitive activity. (1) In the *fault-free (F-free)* state the automatic controller operated perfectly well and no system component failed or was defective. (2) *Practised faults (PracF)* were those fault states that were extensively practised in training. (3) *Novel faults (NovF)* were of the same type as the practised faults but the operator had not experienced them before.

Testing time	t(-2)	t(-1)	t1	t2	t3	....	t55	t56	t57	t(+1)	t(+2)
Mission phase	Pre-M	Pre-M	In-M	In-M	In-M	....	In-M	In-M	In-M	Post-M	Post-M
Cabin crew	A	A	B	B	B	....	B	B	B	A	A

Figure 7. 23: Experimental design (Pre-M = pre- mission phase, In-M = in- mission phase, Post-M = post- mission phase; B = exposure to isolation and confinement, A = non-exposure; within each experimental session: 3 fault states (F-free, PracF, NovF)

Analysis of variance was used to assess changes in performance and subjective state as a function of mission duration and fault type. For that purpose the entire mission was divided into 20 phases (19 in-mission phases plus the post-mission phase), with each in-mission phase consisting of one week. This enabled us to carry out planned comparisons to test some of the specific predictions from the ICE literature, outlined in the introduction. Because of the small number of crew members, the statistical power of the analysis is naturally low. This needs to be considered in the discussion of the findings. Linear regression lines were also used to describe the results of individual crew members. Analysis of variance is often unsuitable for the analysis of N=1 data because of the problem of serial dependency (see Barlow and Herson 1984).

7.4.1.2 Participants

The participants in this study were three Russian research cosmonauts who were selected by the European Space Agency, after having successfully completed the selection procedure of the Mir space programme. Their ages were 31, 35 and 36. The crew members are referred to as CM2, CM3 and CM4 (the back-up crew comprised CMs 1 and 5).

7.4.1.3 Training

After having received an introductory lecture from the experimenter, the crew members were instructed in groups of two or three about the operation of the simulation software. During these software demonstrations, the crew acquired the basic knowledge about the tasks to be completed during the mission. This was followed by practice sessions for individual crew members who received between 5 and 6 hours of training in total. The first practice sessions were administered by the experimenter himself, whereas subsequent training sessions were supervised by a Russian investigator, after he had received extensive training on AMT himself. The training emphasised a system-based training approach, aiming at a deeper understanding of the underlying principles of the system. This was thought to facilitate the crew's comprehension of the effect of control actions on system parameters. This training approach is in contrast to a method emphasising the knowledge of procedures to deal with system disturbances (a more detailed treatment of this may be found in Chapter 6).

As part of the study methodology, crew members were only trained on 6 faults (O_2 leak, O_2 block, CO_2 scrubber ineffective, N_2 set point failure, vent stuck open, cooler failure) out of 16 possible scenarios. This allowed for a distinction to be made between practised faults (predominantly rule-based behaviour) and novel faults (more likely to require real-time problem-solving).

7.4.1.4 Experimental task

In this experiment version 3.0 of AMT was used, which contained a random fault selection facility. This allowed the use of this version in a long-term isolation study that does not permit setting up individual experimental sessions in advance. Crew members were tested three times a week (Mo, We, Fr) for 1 hour. Each AMT session comprised three 20-min periods, with two system faults in each period. The phases before the onset of the first fault and those following the successful repair of each fault were defined as fault-free.

7.4.2 Results

Pre-mission data were incomplete with crew members 2 and 3 having completed only one of the two scheduled testing sessions and CM4 none at all. Because of insufficient data, the pre-mission phase was not included in the analysis. Post-mission sessions were all completed according to schedule.

The small number of participants naturally reduced the power of the statistical tests. In addition, exploratory data analyses revealed that CM4 showed a different pattern from the rest of the crew for most measures. This reduced even further the chance of finding overall group effects. For this reason, a brief description of individual patterns is also given but the data are not presented in detail. All performance and information sampling measures underwent a log-transformation to stabilise variances.

7.4.2.1 Primary task performance

Parameter control failures

The analysis revealed a significant difference between fault types, with NovFs being most difficult to control, followed by PracFs and the fault-free condition (see Figure 7.24). This effect was significant (F=63.6; df=2,42; p<.001). Post-hoc LSD-tests indicated that only the difference between fault-free and each of the two fault types was significant (p<.001). No main effect of mission duration and no interactions were found (both F<1).

The absence of a main effect of mission duration and an interaction was due to pronounced interindividual differences between CMs 2 and 3, on the one hand, and

CM4, on the other hand. Comparing the performance between fault types on each testing day, the results indicated better performance for CMs 2 and 3 on the PracFs than the NovFs during the first mission half (sign-test: CM2: N=28; x=8; p<.05; CM3: N=28; x=5; p<.001). The differences between fault types disappeared in the second mission half. This result represented the interaction that had been expected. No such effect was found for CM4 (sign-test: N=27; x=12; p>.05).

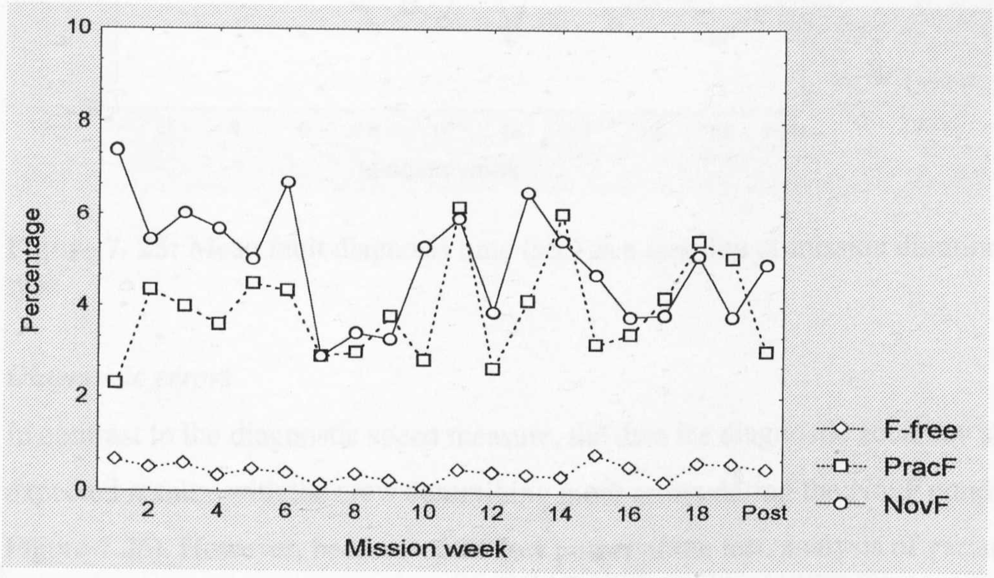


Figure 7. 24: Mean Parameter control failures (%) as a function of mission duration and fault type

Fault identification time

A difference between PracF and NovF in an unpredicted direction was recorded for this measure. Crew members generally needed more time to diagnose a PracF than a NovF (see Figure 7.25). The effect was significant (F=83.1; df=1,2; p<.05). Although an overall increase in diagnostic speed was observed over the mission, the effect failed to reach significance (F=1.47; df=19,38; p<.05). No interaction between mission duration and fault type was found (F<1).

Again, pronounced interindividual differences between crew members were found. CMs 2 and 3 showed a reduction in identification time throughout the mission. A separate ANOVA for these two CMs confirmed this to be highly significant (F=5.42; df=19,38; p<.001). CM4, on the other hand, showed an initial trend of improving performance that was interrupted in the second mission half, with identification times beginning to increase again.

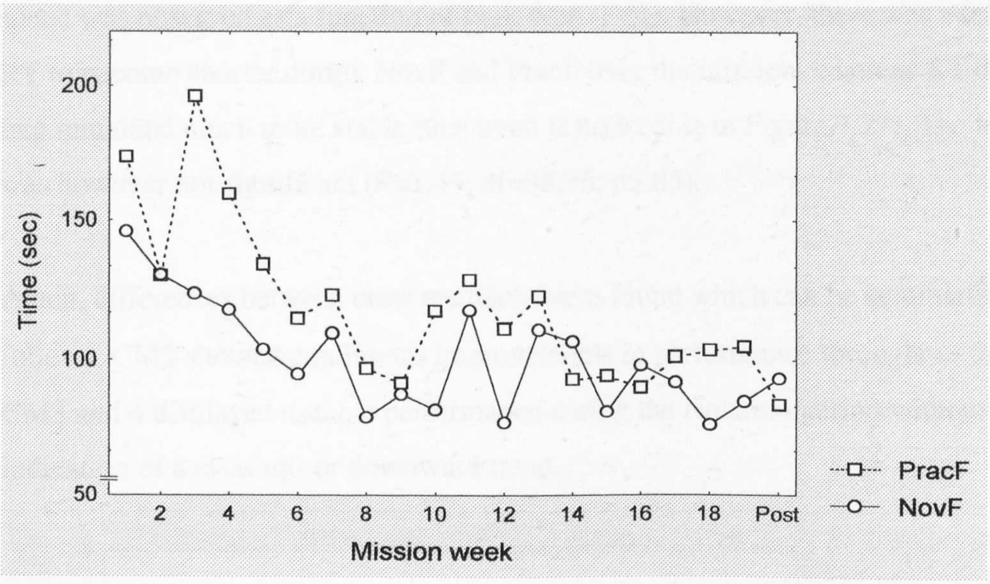


Figure 7. 25: Mean fault diagnosis time (sec) as a function of mission duration and fault type

Diagnostic errors

In contrast to the diagnostic speed measure, the data for diagnostic accuracy showed the expected results, with the crew committing more errors during the NovF condition (see Figure 7.26). However, because of the low power of the test, analysis of variance did not confirm this effect to be significant (F=5.80; df=1,2; p>.05). A simple count of the weeks that showed better performance for practised faults confirmed the difference to be significant (sign-test: N=19; x=3; p<.005). A general reduction in diagnostic errors was observed over the mission but this trend was statistically not significant (F=1.41; df=19,38; p>.05). No interaction was found (F<1).

A different pattern emerged for individual crew members. CMs 2 and 3 showed initial improvements in diagnostic accuracy (over the first few mission weeks), and performance remained reasonably stable thereafter. The data of CM4 showed a rather volatile pattern with peaks of diagnostic errors in mission weeks 3, 4, 14 and 16.

7.4.2.2 Secondary task performance

Alarm reaction time

The data of this measure are presented in Figure 7.27. A fitted regression line indicated an overall reduction in response time over the isolation period. However, analysis of variance failed to confirm the significance of this effect (F<1). No difference in response speed was observed as a function of fault type (F<1). However, there was a tendency for RT to become shorter during NovF and PracF over the mission, whereas RT during F-free remained much more stable (this trend is noticeable in Figure 7.27). The interaction was however not significant (F=1.45; df=38,76; p>.05).

Again, differences between crew members were found which can be summarised as follows. CM2 showed continuous improvements in performance throughout the mission. CM3 and 4 displayed a stable performance during the isolation period without any indication of a clear up- or downward trend.

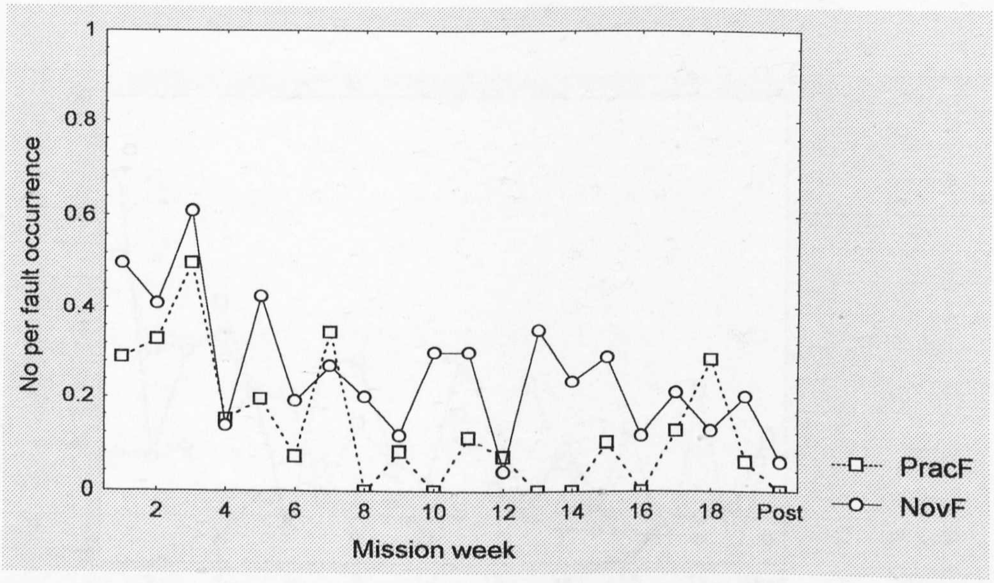


Figure 7. 26: Mean number of diagnostic errors per fault state as a function of mission duration and fault type

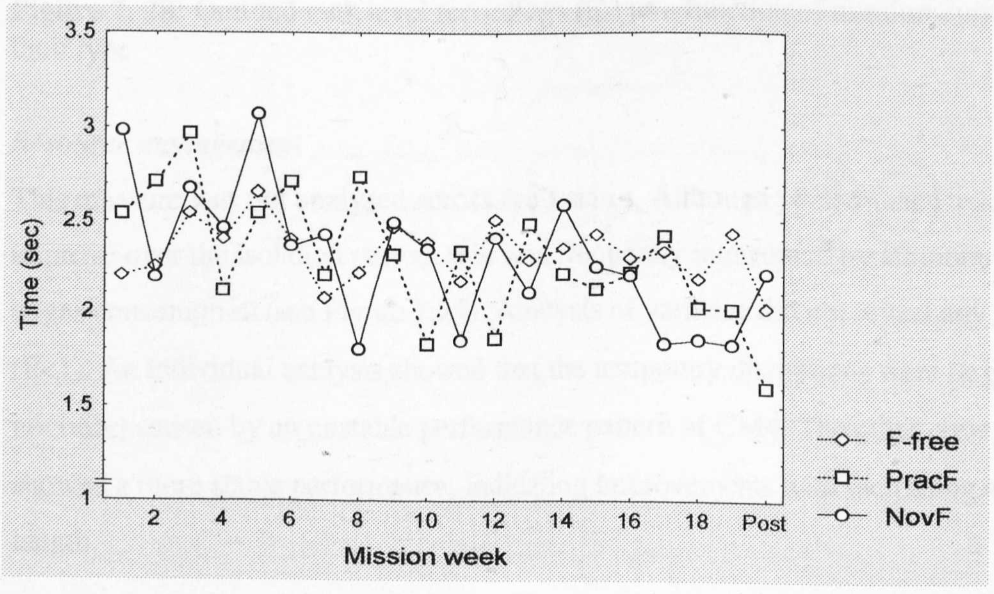


Figure 7. 27: Reaction time (sec) as a function of mission duration and fault type

Prospective memory (tank level recordings)

Crew members were found to omit more recordings during fault states (PracF: 21.4%, NovF: 20.8%) than during fault-free periods (8.5%). This difference was statistically significant (F=11.0; df=2,4; p<.05). Furthermore, the crew showed continuous performance improvements over the isolation period (F=3.29; df=19,38; p<.001). As Figure 7.28 shows, the gap between fault-free and the fault conditions tended to become smaller with increasing mission duration. This resulted in a significant interaction between fault type and mission duration (F=1.60; df=38,76; p<.05). Prospective memory was the only measure that showed a consistent trend for the entire crew.

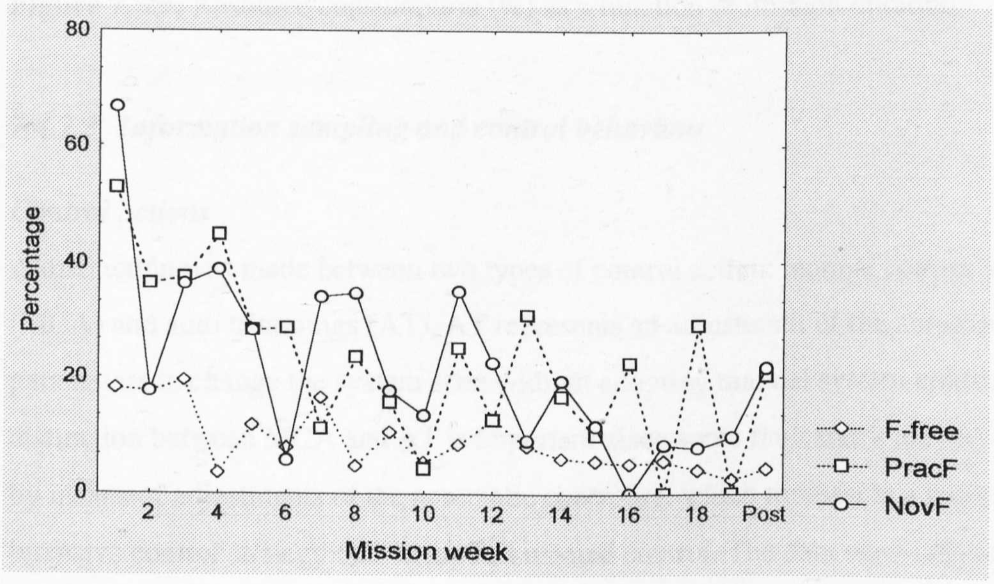


Figure 7. 28: Omitted tank level recordings (%) as a function of mission duration and fault type

Resource management

This measure was not analysed across fault states. Although performance tended to improve over the isolation period, this was frequently interrupted by temporary increases in gas consumption (see Figure 7.29). Analysis of variance did not reveal any clear trend (F<1). An individual analysis showed that the temporary disruptions were largely (but not only) caused by an unstable performance pattern of CM4. The other crew members showed a more stable performance, indicating improvements with increasing mission length.

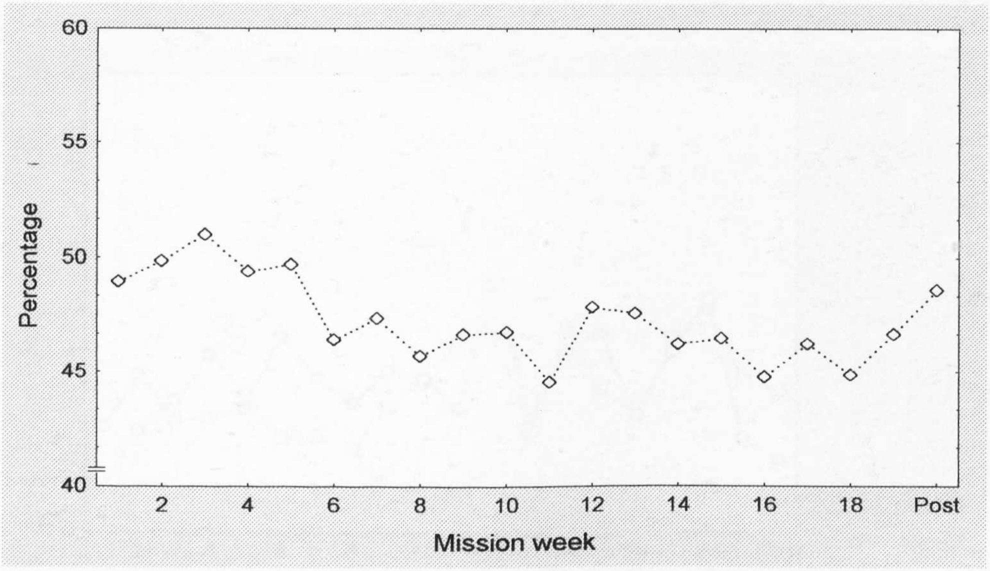


Figure 7. 29: Resource consumption (%) as a function of mission duration

7.4.2.3 Information sampling and control behaviour

Control actions

A distinction was made between two types of control action: manual control actions (MCA) and auto trimmings (AT). AT represents an adjustment of the automatic control parameters to change the system state without adopting manual system control. The distinction between MCA and AT is important since some fault states may be controlled by means of adjustments of the automatic controller, which represents a less resource-intensive control strategy than using full manual control. The data were only analysed between PracF and NovF since virtually no control actions were observed for F-free.

Overall, it emerged that MCAs were carried out more frequently than ATs (1.6 vs 1.1/min). Each control mode is now examined separately. The analysis of MCA did not show any significant difference between fault types (F<1). Visual inspection of the results seemed to indicate that more MCAs were carried out in the later mission phases (see Figure 7.30). However, analysis of variance did not reveal any significant effect (F=1.24; df=19,38; p>.05). No interaction was observed. The data for AT showed that this control mode was used slightly more during PracFs than NovF (1.15 AT/min vs 1.05 AT/min). The difference was however not significant (F=1.75; df=1,2; p>.05). As observed before, this mode of control was also used more frequently with increasing mission duration (see Figure 7.31). However, this tendency was not significant (F<1). No interaction was found.

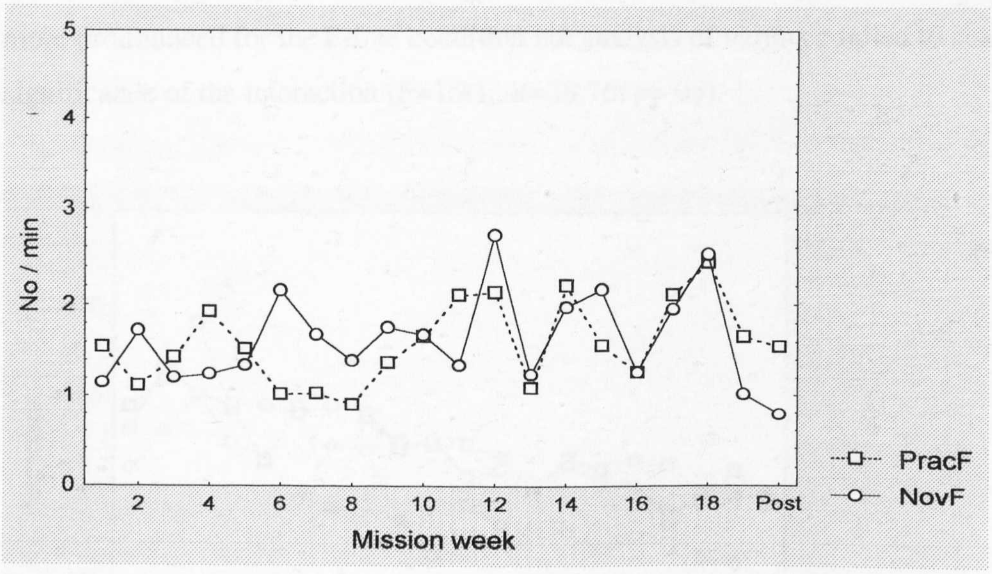


Figure 7. 30: Manual control actions (No / min) as a function of mission duration and fault type

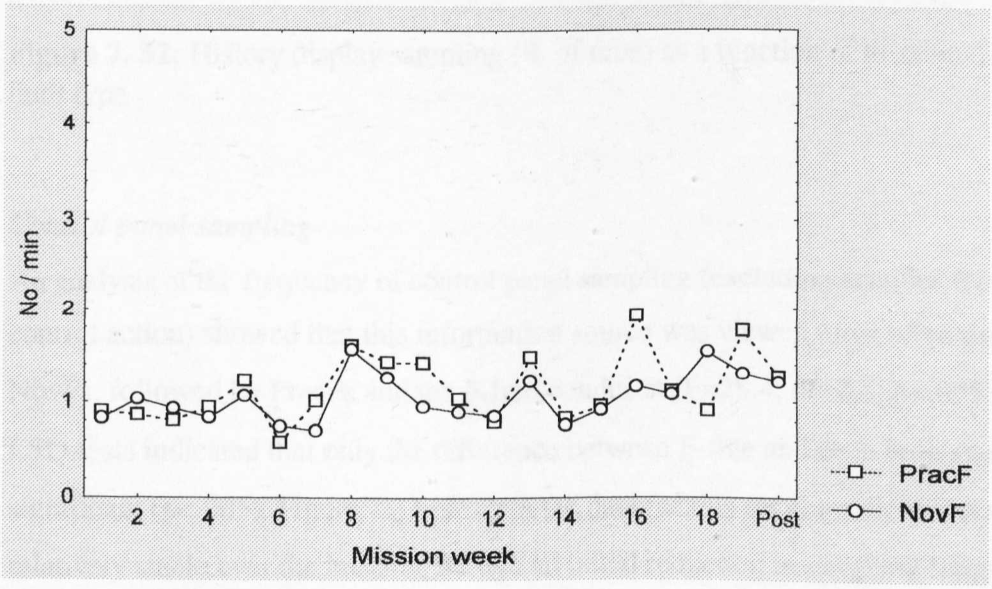


Figure 7. 31: Auto trimmings (No / min) as a function of mission duration and fault type

History display sampling

This measure indicates the total time (as a percentage) the history display was present on screen. When evoked, the display remained on screen for 30 sec. The results showed that it was more frequently used during the presence of a fault (PracF: 50.0%, NovF: 49.1%) than during the F-free condition (37.2%). This difference was significant (F=7.96; df=2,4; p<.05). An overall reduction in monitoring was observed for all fault states during the course of the mission (see Figure 7.32). This virtually linear trend was confirmed to be significant (F=2.28; df=19,38; p<.05). The reduction appeared to be more pronounced for the F-free condition but analysis of variance failed to confirm significance of the interaction (F=1.33; df=38,76; p>.05).

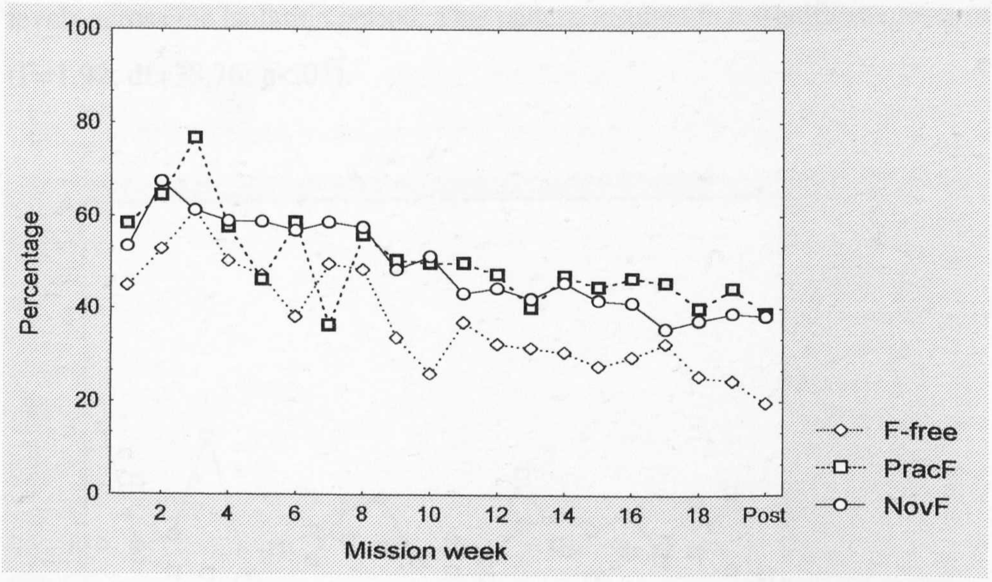


Figure 7. 32: History display sampling (% of time) as a function of mission duration and fault type

Control panel sampling

An analysis of the frequency of control panel sampling (excluding samples followed by a control action) showed that this information source was viewed most often during NovFs, followed by PracFs and the F-free condition (F=28.4; df=2,4; p<.005). However, LSD-tests indicated that only the difference between F-free and each fault state was significant (p<.005). Figure 7.33 presents the data for this measure. Scores remained relatively stable over the mission, though an initial reduction in sampling frequency was observed for F-free over the first few weeks. No effect of mission phase (F=1.66; df=19,38; p>.05) and no interaction were observed (F=1.41; df=38,76; p>.05).

Flow meter sampling

An analysis of this diagnostic sampling measure revealed that fewer samples were taken in the fault-free condition (0.95/min) than during the presence of a fault (PracF: 2.12/min, NovF: 2.16/min). This difference was significant (F=57.6; df=2,4; p<.005). There was also a general trend of reduced sampling activity as the mission progressed but the effect failed to reach significance (F=1.77; df=19,38; p>.05). This was because the reduction in sampling activity was only observed for F-free. During the first few mission weeks, the crew sampled the flow meters with the same frequency in all three conditions, as Figure 7.34 shows. However, as the mission progressed, the sampling frequency for F-free continued to fall, whereas the other two conditions showed stable levels across the isolation period. This pattern resulted in a significant interaction (F=1.92; df=38,76; p<.01).

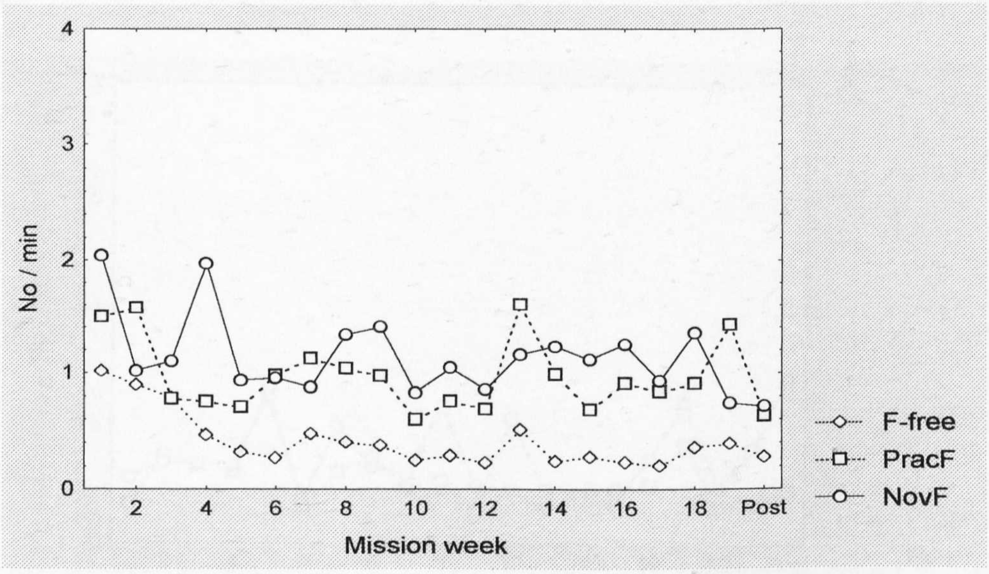


Figure 7. 33: Control panel sampling (No / min) as a function of mission duration and fault type

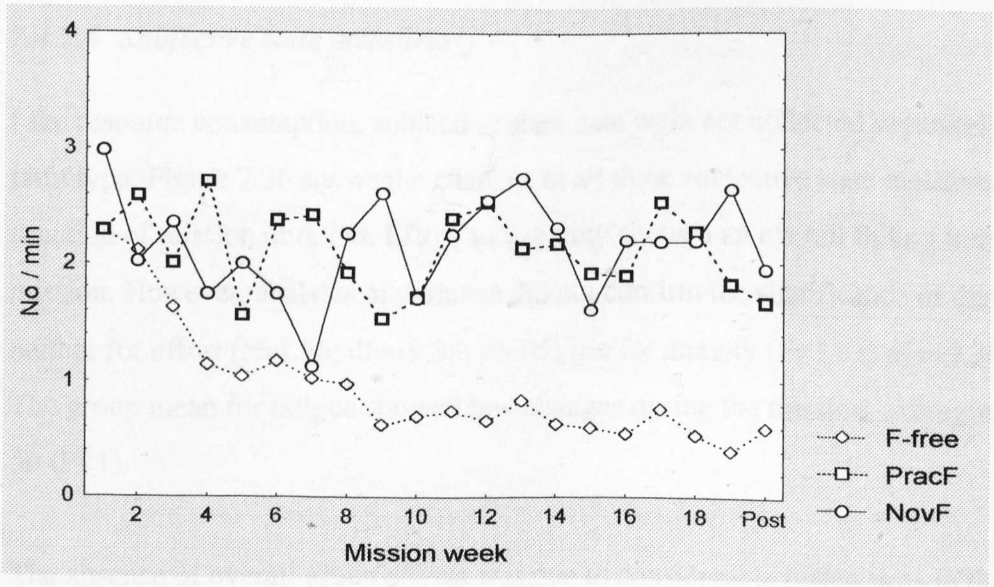


Figure 7. 34: Flow meter sampling (No/min) as a function of mission duration and fault type

Interrogation of maintenance facility

This measure was scored by counting the frequency of accessing the maintenance facility only for inspection (i.e. no repair was carried out). The analysis revealed the following picture (see Figure 7.35). No significant difference was found between fault states (F<1), though NovFs showed a higher sampling rate than PracFs (0.34 vs 0.29 / min). No clear trend was observed over the mission with the measure showing occasional peaks before returning to baseline levels (F=1.17; df=19,38; p>.05). No interaction was found.

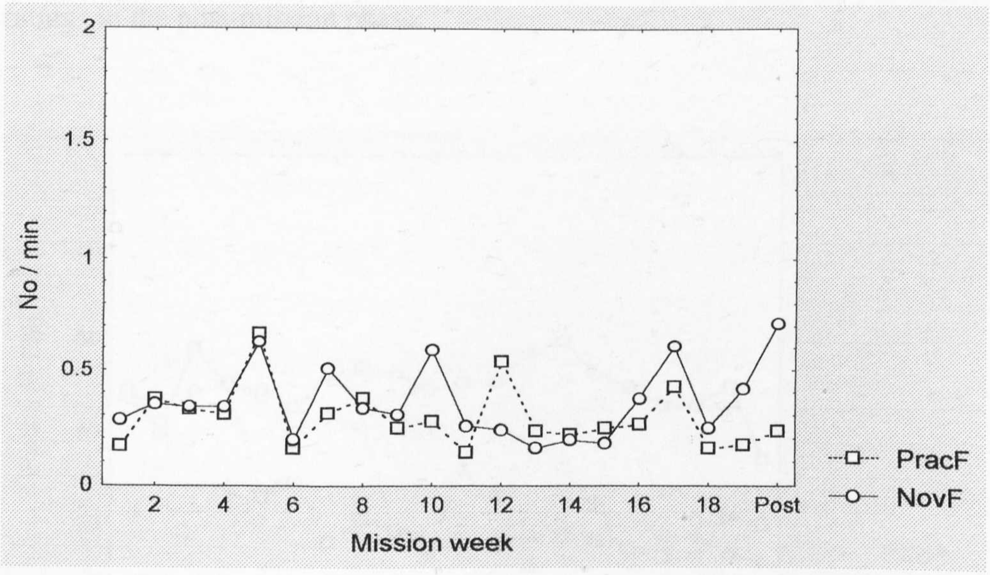


Figure 7. 35: Sampling of maintenance facility (No/min) as a function of mission duration and fault type

7.4.2.4 Subjective state measures

Like resource consumption, subjective state data were not collected separately for each fault type. Figure 7.36 shows the changes in all three subjective state measures as a function of mission duration. Effort and anxiety showed an overall falling trend over the mission. However, analysis of variance did not confirm the significance of this trend, neither for effort (F=1.56; df=19,38; p>.05) nor for anxiety (F=1.81; df=19,38; p>.05). The group mean for fatigue showed few changes during the mission, averaging around 50 (F<1).

The absence of overall group effects was due to considerable differences in the rating pattern of crew members. CMs 2 and 3 showed similar ratings for effort and anxiety. Their ratings displayed a monotonic fall over the first five weeks, followed by stable levels for the rest of the mission. Fatigue levels for CM3 remained stable throughout, with some temporary increases during in-mission weeks 3, 13 and 14. The fatigue ratings for CM2 showed a decrease over the first three weeks and remained at the same level thereafter. Very unusual ratings were made by CM4. Four weeks into the mission, he began to use only the extreme ends of the scales for all three subjective state measures. This resulted in fatigue ratings of close to 100 and anxiety ratings of close to zero. His effort scores began to climb first until they reached 90-100, followed by extreme low ratings in the second mission half. All three state measures showed a return to 'normal' ratings in the post-mission phase.

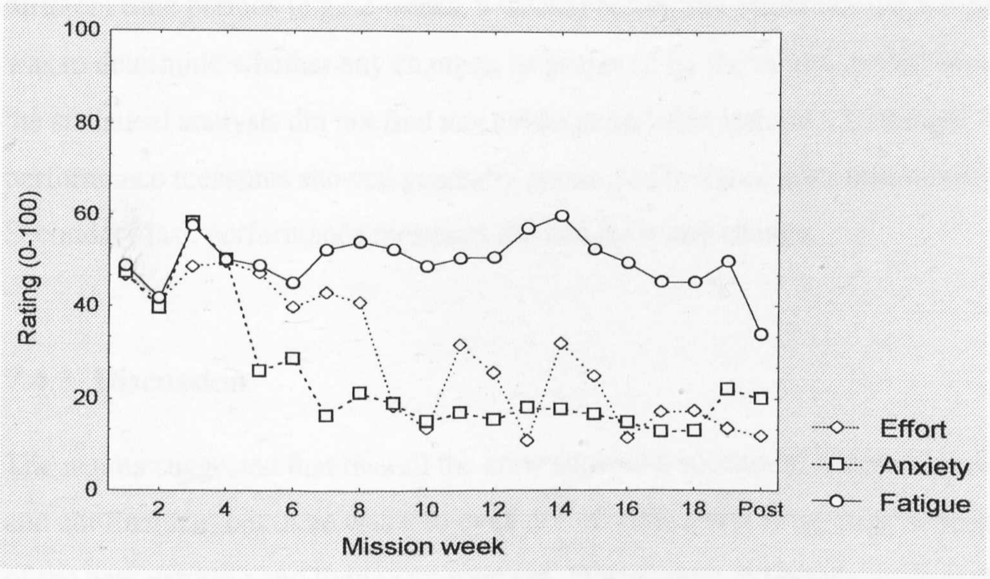


Figure 7. 36: Subjective state measures (0-100) as a function of mission duration

7.4.2.5 In-mission phases

As outlined in the introduction of this chapter (section 7.1.1), some authors proposed the existence of distinct mission phases, which were associated with certain psychological states. Testing these models with our data is considered difficult because of the continuous performance improvements observed during the mission. Despite these problems, an attempt has been made to analyse the available data.

Using planned comparisons within ANOVA was not judged to be suitable since this would not have been a very powerful method to establish effects. Instead, a simple count was carried out, which compared adjacent mission phases for performance and subjective state, separately for each day and crew member.

There was no evidence of a ‘final effort’ effect towards the end of the mission, though performance on all measures (except for diagnostic accuracy) tended to be poorer, as predicted by the model, in the last in-mission week (week 19) compared to the preceding week (week 18). Sign tests on all performance measures did not show significant effects (for all: p>.05).

A second analysis attempted to look at the effect of the mid-mission point by comparing different time periods (e.g., 2 weeks, 4 weeks) before and after mid-mission point. This was to determine whether any changes, as proposed by the model, would occur. Again, the statistical analysis did not find any evidence (all sign test: p>.05), though the primary performance measures showed generally poorer performance after mid-mission point. Secondary task performance measures did not show any change.

7.4.3 Discussion

The results suggested that overall the crew showed a successful adaptation to isolation and confinement but there was also evidence of some strain being experienced as a result of the new working and living environment. Furthermore, it became evident that CM4 showed a fundamentally different adaptation pattern from the other crew members. This resulted in problems of demonstrating the statistical significance of group effects. The power of the statistical tests had already been very low because of the small sample size. As a consequence of the large effect sizes required to obtain significance, it is likely that some type II errors occurred. Nevertheless, a number of significant effects were found and they are now discussed.

7.4.3.1 Primary task performance in ICE

An important question in a study of this kind is whether the strain of ICE impinges upon performance, either in the form of overt decrements or more indirect manifestations. It was hypothesised that possible decrements would be more pronounced for secondary tasks than primary tasks. However, the results did not confirm this hypothesis. All secondary task variables and the two measures of fault diagnosis showed performance improvements during the mission in some form, either confirmed to be a significant effect or, if significance was not reached, visual inspection indicated a trend in this direction. Parameter control failures was the only measure that did not display this trend. How can we explain the absence of any performance improvements on this task? Within the conceptual framework outlined earlier, we need to consider two main opposing effects that impinge upon performance variables. Improvements as a result of continuous practice and deteriorations as a result of strain associated with ICE. In the absence of a control group, it is difficult to determine their respective impact. There are two interpretations of the results. First, the data indicate a hidden decrement (probably as a result of ICE), which is masked by a prolonged continuous learning process. As a result of the two opposing effects, the data did not show a trend in either direction. Second, the learning process for manual control skills was very short, which resulted in stable performance levels being reached very soon. It is argued that the hidden decrement hypothesis is a better explanation for the following reasons. First, the gap between NovF and PracF is only closed after 7 weeks (equals approximately 21 hours of practice). If there had been only little learning involved, the performance gap would have been closed much earlier. Second, judging from the findings from all the experimental work, manual system control is not a trivial undertaking and requires considerable practice to develop a high level of skill. Although this decrement is only of a relative nature (i.e. in contrast to an absolute decrement which shows overt performance deterioration), it may nevertheless be indicative of some strain experienced.

Two interesting questions arise from this. First, why did this effect occur in a primary task measure (which is in contradiction to the methodological framework) and, second, why did it show only in control performance and not in the diagnostic measures? One may suspect that the central assumption that secondary task performance suffers before primary task performance may not apply to ICE as much as to other kind of sub-optimal working conditions. The fundamental difference between ICE and the other stressors researched in this thesis (see Chapters 5 and 6) was time of exposure. It is possible that during long periods of exposure, operators showed a reduced propensity to maintain primary task performance at all costs since this would be more likely to lead to a depletion of their resources. A relative performance decrement, on the other hand, may be acceptable to the operator if overall performance levels are maintained. This has the added benefit of retaining a reserve of resources. The second question is more difficult to answer. One explanation would be that the learning curve is less pronounced for control performance than for diagnostic performance. Although there have been no comparative studies to estimate acquisition functions for these two skills, there might be some indirect evidence. Whereas system control requires a focus of the forward flow of events (what causes what?), the reverse pattern of reasoning (what is caused by what?) is used for diagnostic skills (Landeweerd 1979). Backward reasoning might be more demanding and more difficult than forward reasoning, which would explain the longer learning period found for diagnostic performance. An alternative explanation is that parameter control also contains elements of sensorimotor control, of which the skills may be more easily acquired (and hence require a shorter learning period) than diagnostic skills.

7.4.3.2 Secondary task performance and subjective state in ICE

Concerning secondary task performance, the results did not show any performance deterioration over the mission, with all measures showing some form of improvement, though not always reaching statistical significance. This trend was strongest for prospective memory, which showed continuous performance improvements throughout. Considering the high sensitivity of this measure as a workload indicator, this may be interpreted as a sign of a continuous reduction in overall task demands. However, for one crew member this interpretation is in contradiction to the subjective state rating pattern recorded. The aberrant rating pattern of CM4 is likely to be indicative of considerable strain being experienced about four weeks into the mission. This was supported by group psychology experiments carried out during the isolation period, which also found CM4 to be experiencing considerable strain (Sandal and Berge 1995). A second crew member (CM2) was identified as showing stress symptoms in the form of developing a hyperventilatory breathing pattern during the mission (Wientjes, Veltman and Gaillard 1995). This is a very interesting finding since, judging from our experimental data, CM2 was the best and most stable performer on virtually all tasks, with no indications of stress at all provided by the subjective state data. It may be that CM2 did not wish to report the strain he experienced. This has previously been found in space mission because cosmonauts did not want to endanger their participation in future missions (Gushin 1992). Alternatively, CM2 may not have subjectively felt the strain, which only manifested itself in physiological parameters. This dissociation between different workload measurement techniques is not uncommon since previous research showed some inconsistencies between results obtained with different techniques (Wickens 1992).

Apart from CM4, the crew showed a decrease of fatigue ratings as the mission progressed. This seems to be inconsistent with evidence from space research, where one of the most commonly observed problems was high fatigue levels (Connors et al 1985; Kass 1990; Bluth and Helppie 1987). This discrepancy may be due to the absence of micro-gravity, which may normally contribute substantially to the increase in fatigue. Furthermore, in the present case task-related fatigue was measured, which may be different from fatigue levels experienced at the end of a working day.

7.4.3.3 System management strategies in ICE

For system management the general picture was that the crew reduced system monitoring with increasing mission length whereas system control activities had a tendency to increase. A number of information sampling measures showed significant effects to support this, and the data of most other measures pointed in the same direction. Although no significant effects were found for control behaviour (as a result of the methodological problems referred to above), visual inspection clearly supports this claim. The whole picture suggests that with increasing mission length, the crew has shown a propensity to deal more actively with emerging system disturbances while demanding less information to do it. The reduction of system monitoring may be indicative of a higher risk strategy as a result of prolonged isolation and confinement. Some of the reductions, notably for history display and flow meters, are quite dramatic and may be cause for concern. On the other hand, one may argue that these changes are simply the consequence of selecting a more optimal strategy. Although it is generally difficult to determine the optimal sampling strategy (see Sheridan 1972), there have been indications that the observed changes were not the result of an improved sampling strategy. The F-free condition generally showed the strongest reduction in sampling frequency, which suggests that, with increasing mission duration, the crew adopted a more reactive approach rather than an anticipatory strategy. An anticipatory strategy would consist of regular checking of system components and system state indicators to discover malfunctions at an early stage. The reactive approach, in contrast, implies the use of control actions and system interrogations only after a fault state has been discerned. This change in strategy may also account to some extent for the relative decrement observed for control performance. It may be that ICE has led to a more passive approach to system management, as a result of changes in psychological states (Gushin et al 1993; see section 7.1.1).

It is also noteworthy that increased control activity did not lead to improved control performance (unlike in Experiments 4 and 5). Consistent with Experiments 4 and 5, the data also indicated that the participants did not display a propensity to explore the system in the early mission phases. Exploratory behaviour would have resulted in more control activities at the beginning of the isolation phase, followed by a reduction as the participants became more familiar with the task environment, needing fewer control actions to achieve operational goals.

7.4.3.4 Fault difficulty

The experimental hypotheses predicted better initial performance on PracFs over NovFs, which would progressively disappear with increased practice on the NovFs. There was some evidence of parameter control failures (mainly for CMs 2 and 3) that this prediction would hold, and better performance on the PracFs was observed for diagnostic accuracy throughout the mission. However, none of the other performance measures showed any support for this hypothesis. On the contrary, fault identification times were even longer for PracFs than for NovFs. It is not entirely clear why this occurred but an explanation would be that PracFs were principally more difficult than the NovFs. However, this would be at odds with the results for control performance and diagnostic accuracy, which indicated rather the opposite. A further analysis by individual faults revealed that those PracFs with one of the following two characteristics had longer identification times. (1) They were very easy to control, hence not exerting any pressure onto the operator to make a quick diagnosis since no obvious adverse impact onto the system state occurred. (2) The effect of the fault state on the system was of a non-linear nature with a very subtle impact at the beginning (hence difficult to detect at first), which increased disproportionally over time. There were no equivalent NovFs that showed a similar pattern of accurate diagnosis being correlated with slow diagnostic speed. One may also suspect that crew members may have delayed the diagnostic intervention if they were confident of controlling the impact of the fault state.

The absence of differences between PracFs and NovFs for the secondary tasks suggests that they had similar overall workload levels. This interpretation makes sense because of a fundamental difference between the fault types in the present experiment and in Experiments 1, 4 and 5. Unlike in these experiments, fault types in the present case did not differ in terms of their inherent difficulty but merely in the amount of practice the operator had in dealing with the fault. The difference between fault states and F-free on the most sensitive secondary task (prospective memory) indicated that the presence of a fault increased workload but this increase became smaller as the mission progressed. This suggests that the overall workload demands associated with the task environment decreased over the mission in the same manner as the demands associated with the presence of a fault.

7.4.3.5 Implications

Some words of criticism are deemed to be necessary concerning the set-up of the study. Overall only few effects of mission duration were found in the present experiment. This was partly due to the decision of selecting the system faults at random from a pool of 16 fault states. This random fault selection facility was considered to be beneficial in the light of the requirements for the mission simulation. A very robust task was required since any interventions would be impossible once the mission was under way. Furthermore, the large number of system faults was also expected to increase motivation since the crew would not be faced with the frequent repetition of a small set of faults. However, the large fault pool resulted in a considerable increase in variance because the kind of faults the crew was presented varied from week to week. The increased error variance made it more difficult to obtain significant results. In hindsight, it would have been better to use a small number of faults (e.g. two PracFs, two NovFs) and present one of each type on each testing day, together with four other faults that come from a larger pool. This would have reduced the error variance for PracFs as well as for NovFs while at the same time, sufficient variety would have been maintained by using a large number of faults.

Despite the methodological difficulties encountered in the experiment, it has provided some insight into the adaptation process in ICE but has also raised some new questions. The overall picture was not entirely consistent, with some measures indicating a successful adaptation process (notably the secondary tasks), whereas others hinted at some adjustment difficulties. This pattern supports the use of a broad methodological approach in order not to miss any indications of stress, be it in the form of subjectively experienced strain, signs of physiological maladjustment or decrements in task performance.

7.5 General discussion

The principal issue addressed in this chapter concerns the effects of isolation and confinement. The outcomes of the field research indicated that overall no serious breakdown of performance occurred, though not all participants showed stable patterns of performance throughout. The results also indicated that system control activities tended to increase during the isolation period. This suggests that little system exploration occurred at the beginning of the mission. On the one hand, system exploration is often considered to be of benefit for the development of an accurate and elaborate mental model (see also section 6.1.2). On the other hand, system exploration bears the risk of the system state becoming seriously disturbed (there was some evidence for this in Experiment 1), resulting in poor human-machine-system performance. One would therefore expect that exploratory behaviour is less frequent in real work contexts, in particular if it involves a great risk of causing damage. This may be the reason why the participants in the isolation studies tended to show less exploratory behaviour (i.e. behaving as though in a real work environment) than the student participants. There was some evidence that manual control activity tended to be positively correlated with performance, indicating that higher system control activity was associated with better performance, though the picture was not entirely consistent. There was less evidence for the alternative explanation that manual control actions become increasingly ineffective under sub-optimal conditions, leading to more control actions without attendant performance improvements.

There were indications that prospective memory was more sensitive to increases in demands than the other secondary tasks, though the findings were not entirely consistent. This confirms findings from Experiments 1-3 and is of great importance for the design of multiple-task environments, where the number of secondary tasks is limited (see also section 8.2.2). The picture from the isolation studies did not indicate a general fatigue problem, though some crew members showed increases in fatigue, which tended to occur in the second mission half. This is at odds with data from real spaceflight where the crew frequently reported increasing fatigue during the mission (Kass 1990, Bluth and Helppie 1987). This may be attributed to the absence of micro-gravity, which is known to cause sleep disturbances.

However, there was a major problem associated with the interpretation of the effects of ICE. This was the absence of a control group. This has clearly limited the generalisability of conclusions drawn from the experimental work and has required a more cautious interpretation. We can say that our study participants coped by and large with the strain of ICE but, owing to the lack of a control group, we are unable to assess the precise impact of isolation and confinement on performance, system management behaviour and subjective state. In particular, performance improvements resulting from practice effects may hide performance decrements caused by the strain of ICE. Despite the unsatisfactory situation, one has to recognise the particular nature of spaceflight, which may have to rely more on methodological approaches based on single case studies (as in clinical settings) because of the small participant numbers. This applies to real spaceflight even more than to simulation studies.

The development of a model of mission phases, as proposed by Gushin et al (1993), appears to be a promising endeavour, which helps us to better understand the processes occurring during long-term isolation and confinement. However, in its current form the phase model, it is argued here, is far too unspecific and lacks sufficient empirical data to back it up. In our experimental work there was also insufficient empirical support (mainly based on evidence from Experiment 6) for its predictions about mid-mission point and final mission week. If it is to be applied successfully, it needs to make more accurate and specific predictions about the kind of skills (tracking, diagnostic, manual control, etc.), task management behaviour (information sampling, etc.) and subjective state (including motivational states) affected during certain mission phases. The potential benefits of a good model of this kind are obvious. It would allow us to provide appropriate psychological support during critical periods (see also section 8.5), and task and work schedules could be designed according to the predictions of the model.

In conclusion, one has to stress the importance of using stronger experimental designs than were used (because of constraints) in these isolation and confinement studies and, indeed, in most other studies of this kind. The additional costs associated with improved designs (e.g., sufficient number of testing sessions in time series designs, use of a control group) are often marginal compared to the overall costs of the study. The additional expenditure would disproportionally increase the strength of the design and hence the validity of the data. A further important point is that studies in the Antarctic (such as Experiment 5) are not as valuable as spaceflight simulations with regard to external validity criteria. They differ considerably concerning operational fidelity (Meister 1989; see also section 8.4), notably, because of differences in the ambient environments and the characteristics of the personnel involved. Even though spaceflight simulations represent the best terrestrial analogue of the work environment, it still falls short of addressing the crucial factor of micro-gravity. Micro-gravity is likely to have an effect on some of the dependent variables employed in the experimental work. The implications of this for future space research are discussed in section 8.6.1.

Endnotes:

---

[1] Experiment 6 was conducted as part of the HUBES study.

Chapter 8:

General Discussion and Conclusion

8.0 Summary

The final chapter discusses the theoretical implications for the model of cognitive-energetical control mechanisms. It is concluded that the model is very useful in guiding research and the design of work but may only be of limited use for the prediction of performance or work behaviour because of the high complexity of adjustment processes. A number of variables influencing these adjustment processes are suggested. The utility of a multiple-task environment is evaluated, based on the experimental findings but also on some theoretical deliberations.

Since the scope of AMT is currently limited to skills and abilities required during process control type of activities, a number of suggestions are made concerning the enhancement of the task environment to include other space activities. This includes a number of ecologically relevant secondary tasks, which may prove to be sensitive measures of primary task load. It is also suggested that AMT might be useful, following some modifications, as a selection and training tool. Using the concept of ‘operational fidelity’, AMT is then evaluated concerning its suitability as an experimental tool to advance knowledge in the area it was applied in. It is concluded that the degree of operational fidelity was satisfactory and that some findings can be applied to the ‘real world’ if the limits of the simulation situation are carefully considered.

Based on the results of the experimental work and the literature review, a number of practical recommendations are given for work design in space. For example, the use of technically less advanced systems should be considered, to increase system reliability and facilitate system maintenance because of the self-reliant nature of spaceflight. Furthermore, a number of directions for future research are proposed, which will aid a better understanding of the critical issues in the context of automation and human-machine-interaction.

8.1 Overview of experimental findings

The final chapter attempts to integrate the findings of the experimental work to draw more general conclusions about the theoretical and practical implications of the work. Before these are discussed, the main findings are summarised briefly.

The experimental results suggested that sleep deprivation and noise, as two prominent space stressors, showed rather subtle effects, often only emerging as an interaction with other independent variables (e.g. task difficulty, dialogue control). Where effects were observed, they tended to be more pronounced on non-performance measures (e.g., information sampling behaviour, subjective state), which are not normally directly observable in a work context. Furthermore, in those cases where performance was affected, the effects tended to be restricted to tasks of lower priority. Although there was some degree of commonality between sleep deprivation and noise, clear differences were also found between them. Control activity and effort expenditure increased under sleep deprivation while these changes did not occur under noise exposure. Instead, noise caused increases in anxiety levels. Dialogue control showed very few main effects but a number of interactions with sleep deprivation and task difficulty. This may support the argument that the quality of an interface is not independent of the task and the operator (and even features of the ambient environment), requiring the interface to be tested under specific circumstances.

The experimental work also showed that training is an important factor in manipulating knowledge structures, even though performance measures provided little evidence of differences between training groups. Most evidence for the effects of training stemmed from the information sampling data and the analysis of the mental model. This suggests that the evaluation of training methods needs to go beyond the examination of trainee performance and, as Experiment 3 indicated, an appropriate retention interval needs to be used. The findings also suggested that skill maintenance during extended lay-off periods may be of less concern than commonly thought.

The effects of isolation and confinement were difficult to assess because of the absence of a control group. Although no breakdown of performance was observed in the field experiments, it is not possible to say whether and, to what extent, performance would have been better in the absence of isolation and confinement. Many performance variables showed improvements during the course of the mission, most probably as a result of learning. Few consistent trends were observed for subjective measures, which may be indicative of a more individualised response pattern to isolation and confinement for these variables than for task-related measures. The practical implications of these findings for extended spaceflight are discussed separately in section 8.5 of this chapter.

The independent variables were of two types. The first includes factors that represent sub-optimal working conditions, either in the form of unfavourable operator states (sleep deprivation) or adverse environmental conditions (noise, isolation and confinement). In both cases, the situation in the work environment cannot be easily changed. This is in contrast to the second group, which includes factors such as training and interface, which may directly affect design of work. These variables may be referred to as moderators in this context since they may alleviate the negative effects of sub-optimal working conditions. Although there are some general guidelines available to improve interface design (Bailey 1989) and training (Patrick 1992), the optimal interface design and training method need to be determined in the context of specific task requirements and environmental conditions.

For future experimental work, it may be better to select independent variables from each of the two groups to investigate in what way the negative effects of sub-optimal working conditions may be alleviated by moderator variables. This, of course, is not to say that the combination of two independent variables from one group (e.g., micro-gravity & noise; work shifts & function allocation) would not be worth investigating. To use an analogy from medicine, at first one would usually be more interested in the interaction of a disease with a remedy before investigating the interaction between a number of diseases or a number of remedies. The last two may become more relevant in the advanced stages of a research programme.

Workload as a variable cannot be assigned to either group since it represents a rather special case. While it clearly does not represent a sub-optimal working condition, neither does it correspond to a typical moderator variable (in the sense of the definition given above). Let us look at the following example. A study is conducted, using interface as the moderator variable. If the results were such that, under certain sub-optimal working conditions, performance had generally been better with ‘interface A’ than with ‘interface B’, we would have chosen ‘interface A’ (for reasons of simplicity, we will ignore all other factors contributing to the quality of an interface). If we had used workload as an independent variable instead of interface, we would not have been able to draw the same conclusion. If it had shown that performance was better under lower workload levels than under higher workload levels, it would not have made sense to give the recommendation that workload levels should be reduced (unless we tested a specific system for the maximum workload the operator can handle).

Workload is different in that it is a critical factor in producing effects of the independent variables. The experimental work showed that only under appropriate workload levels, effects of the independent variables were observed. Therefore, it is important, but sometimes difficult, to select appropriate workload levels, which simulate normal system operations as well as disturbed system states of different kinds, ranging from mild disturbances to catastrophic breakdowns. If the range of levels covered in the different experimental conditions is set too narrow at the lower end of the workload continuum, no effects will be observed and therefore, the impact of the independent variable will be underestimated. On the other hand, if levels are set too high, the task will become too difficult to manage and the findings may only be relevant to emergency situations.

Having discussed some issues surrounding sub-optimal working conditions and moderator variables, an attempt is now made to explore the specific effects of the independent variables in relation to the model of cognitive-energetical control mechanisms.

8.2 Theoretical implications

8.2.1 Model of cognitive-energetical control mechanisms

8.2.1.1 Prediction of adjustment patterns

One of the central theories guiding the research for this thesis was the model of cognitive-energetical control mechanisms (e.g., Hockey 1993). Based on the empirical outcomes of the research, the model is now critically revisited. The model suggests a number of adjustment patterns of how operators may cope with excessive work demands. We need to ask the question of how difficult it would be to predict the adjustment pattern that was to occur under given circumstances? One would expect that this would not be an easy task because of the many factors which may determine the selection of one adjustment pattern for another. A number of factors that may influence this selection decision are now discussed.

(1) The *costs of a performance breakdown* are an important factor since it will determine the extent to which the individual will attempt to protect primary task performance. These costs may be in the form of a financial loss, injury or death of the operator or of other people, or serious environmental damage. (2) There are also differences between work environment concerning the availability of *less critical aspects of work.* These may be neglected or sacrificed without endangering the maintenance of primary task performance in the event of high task demands. Jobs differ widely with regard to the possibilities they offer for neglecting less critical aspects of work. (3) One may expect that with longer *exposure* to the stressor, there will be a reduction in the use of increased effort and sympathetic activation to maintain task performance. This follows from the assumption in resource theory that the resource pool has only a limited quantity available for expenditure. (4) The *mood* or the *emotional state* of the individual may determine the likelihood of using a low effort strategy or of being prepared to adopt a riskier strategy. For example, Shingledecker and Holding (1974) found that fatigued participants adopted a riskier decision-making strategy, and Maule and Hockey (1996) provided evidence that mood affects risk-taking behaviour, though the relationship may be quite complex. (5) *Commitment* to task or organisational goals is also expected to be positively correlated with performance maintenance. (6) Closely related to this may be the *motivation* of the individual to perform within acceptable performance standards. This would include interindividual as well as intraindividual variations in motivation. For example, individuals with high achievement motivation may be more likely to increase effort expenditure to maintain performance, or attempt to maintain it for longer, than those with low motivation. (7) If performance is not easily attributable to individuals, they are more likely to show performance decrements. This effect has been called social loafing (e.g. Franzoi 1996). This may occur, for instance, in a group situation, where the contribution of each group member to overall performance is not readily identifiable.

This list of factors is not meant to be exhaustive but aims to illustrate the complexity of the problem of predicting adjustment strategies used by individuals. It may be possible to make quite accurate predictions for single factors but when including interactions, this task becomes extremely difficult if not impossible. These factors also include concepts that are not typically researched in the context of human factors but rather belong to the wider field of occupational psychology.

A few qualifications also need to be made about the distinction between primary and secondary tasks. This is an a priori distinction made by the experimenter in the laboratory or by the organisation in the context of real work. However, priorities may not always be pursued as required for a number of reasons. First, in some work settings it may not have been clearly conveyed what priorities are attached to the different tasks. Second, priorities are often externally imposed and not internally accepted by all individuals. Therefore, individuals may be of a different opinion than management of what the priorities are in a job. Third, the individual does not feel sufficiently confident to do the priority task because he/she does not possess the appropriate skills, and hence will neglect it (this may be possible in work environments where the pursuit of primary task goals is poorly monitored by the organisation). Fourth, priorities are not adhered to because the priority task is associated, for instance, with too much danger, boredom or difficult hygienic conditions. In order to identify the individual prioritisation of work tasks, a more detailed analysis is required. This also applies to laboratory tasks. If priorities are not followed as expected, the results cannot be easily interpreted within the framework of the model. As for studies within the 'micro-world' paradigm, it is therefore important to provide a good rationale for the prioritisation of work tasks (see also section 8.3).

The model makes little explicit reference to whether these response patterns reflect an intentional decision of the individual or are the product of an automatic adjustment pattern. However, in one instance attentional narrowing is considered to be 'a strategic adjustment rather than an automatic consequence of stress' (Hockey 1993, p333). The general emphasis is on the consciousness of adjustment processes, which is also reflected in the underlying principles of action theory (Frese and Zapf 1994). However, it is argued here that it is equivocal whether operators are always aware of the strategies they use. While the maintenance of primary task performance by means of increased effort expenditure represents most probably a conscious decision, the degradation of secondary task performance as a result of attentional narrowing may not necessarily. This is particularly likely if information sampling has become an automatic process, which is executed at the skill level and does not require much higher-order control (Rasmussen

1986; Reason 1990). Action theory also acknowledges that activities at the sensorimotor level of regulation may take place without conscious attention (Frese and Zapf 1994). The question of whether the response pattern is intentional (and conscious) or automatic (and possibly unconscious) is not only of theoretical relevance but also of high practical value. If these strategies represented conscious actions, they would be part of a range of means at the disposal of the individual to deal actively with changes in work demand. If they were not, this would represent an increased risk for operators with their performance levels degrading without their being aware of it.

8.2.1.2 Occurrence of adjustment patterns in experimental work

Based on the results from the experimental work, an overall evaluation is now carried out to determine the occurrence of specific adjustment patterns.

The protection of primary task performance as a main goal was generally observed in the experimental studies. However, in the case of isolation and confinement this picture was not entirely corroborated. There were some indications that the protected nature of primary task performance was partly disabled. It is however not clear why this happened. It might have been caused by the length of exposure to this stressor. In all other experiments, exposure to the space-related stressors was only short or medium-term. Alternatively, the decrement may have been the direct result of reduced monitoring activities, or it may have been due to a shift in priorities from system control activities to fault diagnosis. Nevertheless, the empirical data generally confirmed the postulate of the model that the maintenance of primary task performance is given priority.

Augmenting effort expenditure seems to be used as a strategy only if the increase in demands is of a short-term nature. The probable duration of the high demand situation will largely determine the amount of mental resources required and the operator will probably wish to avoid a depletion of his/her mental resource pool. Therefore, this represents a well-adapted strategy since increased activation over prolonged periods is generally considered as counterproductive, resulting in increased errors and reduced error detection rates (Frese and Zapf 1994). The use of increased mental effort may also be contingent upon the absence of fatigue. One may also assume that effort is regulated according to the importance of task goals so that during real emergencies extra energetical resources are made available. This corresponds to the biological fight or flight response that also releases additional energetical resources, which would not be available in a non-critical event.

During excessive task demands, the operator tends to opt primarily for a reduction in monitoring activity. In terms of the model this corresponds to the adjustment pattern of increased risk-taking / effort reduction. In many ways this appears to be a good strategy because it does not necessarily lead to a decrement in performance. One takes a calculated risk to maintain performance by expending less mental resources on monitoring while hoping that no critical system event will be missed. An alternative strategy, the neglect of a secondary task, is much more likely to lead to a performance deterioration. One may consider these adaptive changes as a shift from proactive towards a more reactive approach. The difference between a reactive and proactive approach is similar to Hacker's (1986) distinction between *momentary* and *planning strategies*. The latter involves a longer time frame for forward-planning and the anticipation of errors and was found to result in better performance.

There was no evidence of a reduction of control activities during sub-optimal working conditions. One might have expected control activity to decrease in a similar way to system monitoring measures to conserve resources during periods of high demand. However, the overall picture rather indicated the opposite trend, with control activities often showing a tendency to increase. However, the reasons for this were not unequivocal. On the one hand, there was evidence for manual system control becoming less effective and hence, more control actions were needed. On the other hand, there was also evidence for the opposite pattern with an increase in control activity being associated with improved performance.

In conclusion, the model of cognitive-energetical control mechanisms has proved useful in guiding the research in this thesis. Because of the complexity of the adaptive processes and the many variables influencing these processes, it is very difficult to predict an individual adjustment pattern to excessive task demands in a given situation. The broad research approach was perhaps ill-suited to refine the model since too many variables may have influenced the process. Many of these variables were neither controlled nor measured. In order to develop the model further, one needs to conduct more tightly controlled experiments, which allow the testing of some hypotheses. If the factors with the highest predictive power are identified, one can then proceed to simulation studies in settings with higher ecological validity.

8.2.1.3 The broader viewpoint of system design

The adjustment patterns suggested by Hockey (1993) represent useful adaptive strategies from the individual's point of view, though these are not necessarily desirable from the organisation's point of view. An attempt is now made to apply the model in a wider context by considering it from the perspective of work design, where certain adjustment patterns observed among individuals need to be avoided for reasons of overall system effectiveness. Not only do these issues apply to space travel, but they are also of relevance in many terrestrial work environments.

Some of the adjustment patterns proposed appear more critical than others for the organisation. Table 8.1 summarises the patterns of decrement and their respective evaluation of the consequences[1]. The adjustment patterns are now discussed in turn.

Adjustment pattern	*Positive consequences for organisation*	*Possible negative consequences for organisation*	*Criticality*
Primary performance breakdown	None	Collapse of total system performance, accidents, loss of production	***
Sustained high effort / increased sympathetic activation	All aspects of task performance are maintained	Risk of stress-related diseases, absence from work because of illness	*
Lowering of work goals	Performance breakdown is avoided	Reduction in production quality and production quantity	**
Increased risk-taking / low effort	Performance breakdown is avoided	Risk of reduced output, accident	**
Neglect of secondary tasks	Central aspect of task performance is maintained	Reduction in work quality	**

Table 8. 1: Patterns of decrement with positive and negative consequences for organisation, and assessment of criticality of negative consequences for organisation (the more stars, the more critical)

The worst outcome of excessive work demands from an organisational point of view is a complete breakdown of human performance. Depending on the nature of the work, this may result in major accidents (aircraft crash, radioactive contamination, etc.) and / or excessive costs for the organisation (loss of production, damage to machinery, etc.). This kind of outcome is least desirable because it may be critical to organisational success or may even threaten organisational survival.

The second pattern (sympathetic activation / high effort) is often considered less critical by organisations since the consequences are often taken by the individual rather than the organisation. Prolonged high effort expenditure and / or increased sympathetic activation may result in stress-related health problems. These may have ramifications in the short-term (e.g. absence from work owing to illness) as well as in the long-term (e.g. early retirement because of work-related diseases). A number of organisations use this as a deliberate strategy to survive in a very competitive market in that they exploit their employees as far as possible and do not take responsibility for the long-term costs of ill health (Morgan 1986). Costs for medical treatment, early retirement, etc. are paid for by society rather than being charged to the organisation that caused it in the first place. One may even suspect that in some cases work is designed such that the strategies at the disposal of the operator are deliberately reduced by means of work design (e.g. machine-paced assembly line work).

The lowering of work goals may also have undesirable effects for the organisation in that the staff reduces the work pace, if possible, or the quality of work is compromised, or both. However, this is not as critical as a complete performance breakdown. The remaining two adjustment patterns, increased risk-taking / low effort expenditure and neglect of secondary tasks, are considered to have a similar level of criticality. Both may result in additional costs or lost revenue for the organisation but not on the same scale as a complete performance breakdown. Which one(s) will be chosen depends on a number of factors, such as the nature of work since this often limits the options at the disposal of the individual to manage varying task demands. Some of these factors, which may influence the individual to opt for a particular adjustment pattern, are discussed below.

It is not argued here that these adjustment patterns represent ill-suited strategies that are counterproductive. On the contrary, they represent well adapted strategies on the part of the individual to cope with changing demands at work, which may, at times, exceed available resources. However, it is argued that a goal of work design should be to minimise these excesses in demand since the adaptive strategies may have negative side-effects for the individual (e.g. ill health) as well as for the organisation (e.g. reduced work quality or quantity). A further goal is to reduce the negative consequences of excessive work demands by designing work such that non-central aspects of work can be (1) temporally postponed to a less busy period, (2) delegated to a colleague, or (3) delegated to a machine. These strategies would correspond to 'neglect of secondary task performance' without any negative consequence for work quality because the task is completed at a later time or by a third party (human or machine). Work design may also incorporate the possibility of being able to control the pace of work, which would allow the individual to slow down without necessarily sacrificing quality of performance. All measures suggested should provide more flexibility in the design of work to enable the individual to cope better with variations in work demands.

8.2.2 Multiple-task environments

The evaluation of the model of cognitive-energetical control mechanisms required the use of a multiple-task environment to be able to investigate the different adjustment patterns. The usefulness of multiple-task environments is now examined.

Most multiple-task environments used in previous studies are probably more accurately described as dual task environments since few have used more than two tasks (see Proctor and Dutta 1995; Wickens 1992). Many of the simulation tasks reviewed in Chapter 4 (see section 4.1.3) represent only single or dual-task environments, though of considerable complexity in some instances.

A number of problems are associated with the use of multiple-task environments as opposed to dual-task environments.[2] Because of the higher complexity of multiple-task environments, it is more difficult to determine the amount of resources that are allocated to each task. Furthermore, this may provide the operator with the possibility of rescheduling a task rather than resorting to time-sharing. The likelihood that experimental participants disobey instructions and violate task priorities (see section 8.2.1.1 for an account of possible reasons) is probably greater in multiple-task environments. For these reasons, it might be difficult to determine an accurate performance-resource function, though this may be possible mathematically[3].

Despite the problems associated with the use of multiple-task environments to determine accurate performance-resource functions, it is argued that, nevertheless, they represent a useful approach. Multiple-task environments may be less suitable for determining accurate performance-resource functions but they have other advantages. Since they allow the measurement of a number of different skills (e.g., diagnostic performance and manual control performance in the present case), it is possible to detect differential effects of the sub-optimal working conditions on a range of skills. The experimental results demonstrated this on a number of occasions. Operational fidelity is also higher in a multiple-task environment because it does not eliminate some tasks for the sake of more stringent control (see section 8.4 for a more detailed discussion). With user acceptance and user motivation being closely related to operational fidelity, these factors may gain critical importance in applied settings.

The issue of specific task interference is also of relevance here. Within the AMT environment, there was a strong similarity between tasks concerning input modalities and output modes. All tasks used the visual modality to encode information while manual responses were required to communicate with the system. Since intra-modal time-sharing is considered to be more vulnerable to interference than cross-modal time-sharing (Wickens, Sandry and Vidulich 1983; Parkes and Coleman 1990), we may have had some specific task interference within AMT for input modalities and output modes. However, one may argue that a certain degree of task interference is desirable since it increases the sensibility of the simulation task. This allows it to be more sensitive to sub-optimal working conditions.

Examining the vulnerability of the secondary tasks in AMT, the data clearly suggested that prospective memory was a more vulnerable task than reaction time, across the whole range of independent variables used in the studies. It is not clear whether this is because fewer specific resources may be available for prospective memory, or the RT task may be more data limited. This finding is independent of any possible specific interference between prospective memory and any of the primary tasks since no attempt was made to compare the effect of primary task completion on prospective memory. Instead, we were interested in the impact of the independent variables on primary and secondary tasks. The finding of prospective memory representing a more sensitive task may be valuable information for future studies that are limited in the number of secondary tasks they can use. Indeed, a study evaluating the sensitivity of various measures of workload found that the most sensitive measure was time estimation (Casali and Wierwille 1984), which shares some elements with prospective memory. The third secondary task 'resource conservation' was different from the others in that it was closely related to system control performance (a primary task), but emphasised the efficiency of control, which is not identical to the primary task goal of effectiveness. The reason why this measure showed few effects may be that it was treated like a primary task because of its close connection with system control performance. Another possibility is that alternative control strategies - which should have led to the same goal (system remains in target state) but by different means (high vs low resource intensive strategies) - were not sufficiently available or not apparent to the operator.

When selecting suitable tasks, in particular secondary tasks, for research of this kind, further points are to be considered. These tasks need to fulfil some basic requirements. First, they need to be of some ecological relevance in the work environment. If not, they run the risk of not receiving sufficient user acceptance, which may influence work behaviour considerably. In the absence of sufficient ecological relevance, one can improve acceptance by providing an acceptable 'cover story' (Brehmer et al 1991). The issue of ecological relevance is highly important since it may also have the opposite but, nevertheless, undesirable effect of a secondary task being treated as a primary task because it is considered too important. This is the main problem associated with the loading task (Ogden et al 1979), where the respective roles of the primary task and the secondary task are reversed with the original primary task now being treated as a secondary task. Although experimental participants generally follow the instructions of the experimenter, they may implicitly alter the weights assigned to each task because they feel it contradicts natural task priorities (see also discussion in section 8.3). This problem however is greater in a study with high operational fidelity (where unnatural task priorities would stand out) than in a lab-based study that makes use of tasks with little ecological meaning.

8.3 Enhancement of AMT

Based on the discussion of some of the difficulties in developing an appropriate multiple-task environment, some directions for enhancement of the experimental task are proposed. Furthermore, this section explores some potential applications for AMT, which go beyond its current use as a tool for human factors research.

The suggestions to enhance AMT refer to primary and secondary tasks, though the allocation of priorities may depend on operational conditions. With AMT currently based on the life support system of a spacecraft, the task environment could easily be expanded by incorporating further technical system used in a spacecraft, such as propulsion, communication and even payload facilities. This would allow the measurement of crew skills that are not assessed with the current AMT versions but were identified as relevant in the task analysis in Chapter 1. Table 8.2 summarises the proposed tasks and their suggested task priority.

	Primary task	*Either*	*Secondary task*
Tracking (spacecraft docking)	X		
Verbal reasoning		X	
Numerical reasoning		X	
Spatial reasoning		X	
Working memory			X
Short-term memory			X
Monitoring (delegation task)			X

Table 8. 2: Proposed tasks for AMT enhancement

As part of the propulsion system, a *tracking task* may be used in the operational context of docking manoeuvres. During communication with other groups, different kind of information such as verbal and numerical material needs to be understood and interpreted. This task could be modelled on psychometric tests for *verbal* and *numerical reasoning*. In addition, it may include *spatial reasoning*. The kind of information presented in these tests needs to be adapted to make it more relevant to spaceflight. The items could be displayed as 'email messages' requiring a decision based on the information presented. One may expect these reasoning tasks to be very resource demanding, which means that sub-optimal operational states are likely to become manifest.

A task that places high demands on *working memory* capacity is also proposed to provide an indicator of monitoring load. This may be in the form of a contaminant monitoring task, which requires the operator to check the safe cabin levels of space contaminants. The task could represent a simplified version of COST, which has already been used in space research (Hockey and Sauer 1996). It may involve working with verbal material, numerical information, or both. One could also include a *short-term memory task* involving the temporary storage of names and positions of fictitious planets during cosmic observations. An interesting indicator of excessive workload would be a *'call crew member'* button, which would effectively 'delegate' a task to another crew member (in this case the computer) while the operator would be temporarily relieved from this task until he/she had sufficient spare capacity to take it on again. This task represents a modification of a task feature in a ship's bridge control task (Kerstholt, Passenier, Houttuin and Schuffel 1996), which included a 'global reset' button, allowing the watchkeeper to reduce his/her workload temporarily while dealing with a system fault.

As a result of the complexity of the proposed task environment, the design needs to follow a modular principle, which allows for the selection of those modules that meet the requirements of the specific situation. The increased complexity of the task environment has two important implications. First, it will lengthen training time. This may be a critical factor during simulated or real spaceflight when allocated training time is at a premium. Second, it will also lead to an increase in testing time since each module requires a minimum testing period to obtain a meaningful baseline. Again, during some missions little time may be available for extensive testing, though other simulated or real flights require the provision of meaningful crew activities during slack periods (e.g. cruise phase). A possible third implication of increased complexity might be that not all sub-tasks in the task environment can be completed concurrently. One may have to allocate time slots for the completion of certain tasks according to the single-task principle. This means that the main process of AMT is temporarily interrupted and recommences as soon as the other task has been completed. This would introduce a sequential element into a task environment that is otherwise characterised by concurrent task elements.

The availability of a complex task in a modular form would have the advantage of being able to manipulate the task environment on a number of dimensions. First, *task difficulty* may be varied by manipulating the number of decision aids (history display etc.) available to the operator during the completion of each task. Second, *task load* may be varied by manipulating the number of separate tasks (e.g., manoeuvring, communication, life support) used in the overall AMT environment.

In addition to using AMT for generic human factors research, there are two more potential areas of application. First, AMT may be used as a selection tool to test candidates on some skills for their suitability for becoming an astronaut. This would naturally require a number of validation studies to establish the psychometric properties of AMT. Although dynamic tasks are rarely used in selection, it may have the advantage of achieving higher predictive validity scores than an item-based test. Second, AMT could be used in a training context. However, for this purpose it would need to be made more complex to ensure that even after extensive training, the task remains sufficiently challenging.

8.4 Fidelity of the simulation

The methodological approach adopted in this thesis was predominantly influenced by the 'micro-world' paradigm (Brehmer 1992, see also Chapter 4). The question is, of course, to what extent the findings obtained with a computer simulation task can be transferred to the control of a life support system during real spaceflight - in Meister's (1989) terms, the 'reference situation'.

A number of important differences are noticeable between the reference situation and the non-reference situation. The size of these differences determines the degree of operational fidelity. One may distinguish between different dimensions of operational fidelity: *environment, operator task, human-machine-interface* and *operator characteristics*. These are derived from the 'cognitive triad' (Roth and Woods 1988; see also Chapter 3) and Meister's (1989) criteria (see Table 8.3).

Experiment	Dimension			
	Environment	*Operator task*	*H-M-Interface*	*Operator*
1 (sleep, control)	low	upper medium	medium	medium
2, 3 (training)	low	upper medium	medium	medium
4 (6-day ICE)	high	lower medium	medium	high
5 (8-month ICE)	medium	lower medium	medium	lower medium
6 (135-day ICE)	high	upper medium	medium	high
Overall	*low to high*	*medium*	*medium*	*medium to high*

Table 8. 3: Degree of operational fidelity of experiments as a function of fidelity dimensions

The degree of fidelity of the ambient space environment varied considerably between experiments. It was quite high in the two spaceflight simulation studies (Experiments 2 and 5), where only micro-gravity and mission danger were absent as features of the ambient space environment. In particular, the first is considered as very important (see also for the discussion below) since it has a considerable influence on performance (Manzey et al 1995, Connors et al 1985). Mission danger is probably a less important factor since, providing the mission runs smoothly, one can expect the crew to adapt quickly to the new working and living environment, with a reduction in perceived environmental threat. However, in the event of an emergency, high anxiety may result in similar patterns of behaviour found in terrestrial emergency situations such as attentional narrowing (Baddeley 1972), impairment of working memory (Davies and Parasuraman 1982) and non-optimal shifts in speed-accuracy trade-off (Wickens 1992). The lab-based experiments, on the other hand, were situated at the lower end of the fidelity continuum since they lacked a number of features of the ambient space environment (e.g. social isolation, hygienic conditions, extended task involvement, confinement).

The task environment (i.e. the H-M Interface) may be considered as a medium fidelity simulation. It represents a generic life support system, which is modelled on a relational scale rather than a physical scale, that is, there are no detailed representations of the physical layout of system components. However, this is not necessarily considered to be a disadvantage, providing the simulation models the essential features of the reference system (Meister 1989). In a real system the emphasis is more on monitoring and checking while manual control and diagnostic activities are less frequent than in the simulation. This is because the process of the simulation needs to be speeded up and the frequency of system failures needs to be increased, in order to collect sufficient data within a testing period of acceptable duration.

The task itself may be considered as being medium fidelity, though there were slight differences among versions concerning the complexity of the task (see Chapter 4), hence the differentiation. Its complexity in absolute terms was lower compared to a real system. However, in terms of relative complexity it was certainly comparable since this is primarily a function of training and experience. Judged by the difficulties the operators experienced in managing the system effectively, its relative complexity appeared to have been more than sufficient. Furthermore, one cannot evaluate the system's characteristics against criteria of good system design since we developed the system for the purpose of measuring performance and system management behaviour rather than aiming to develop a system that would allow optimum control of the life support system. Hence there are always components and tasks that appear to be rather 'artificial'. For example, the prospective memory task represented poor system design (no prompt for the operator) but it fulfilled its purpose of measuring primary task load.

The degree of operator fidelity (or personnel fidelity, as Meister calls it) ranged from medium (for the postgraduate student participants and the Antarctic over-wintering crew) to high (for the trainee astronauts and cosmonauts). Because of the selection criteria used for the recruitment of the student participants, one can assume that they were not too dissimilar to astronauts in terms of ability, though the latter group is likely to have developed a higher skill level as a result of many years of training. With regard to motivation, one can assume that students are sufficiently motivated (though probably not as much as a real space crew) for a testing session of a few hours. The motivational problem may be greater during an isolation and confinement study but one may also suspect that similar effects are found in real spaceflight, where the crew may also have difficulties in sustaining motivation for routine activities.

When we take into account all the facets of fidelity, we may describe the operational fidelity of the non-reference situation, on average, as medium. This leaves us now with the question of the validity of any conclusions and recommendations derived from the non-reference situation and generalised to the reference situation. Since the research was generic and did not attempt to evaluate any particular system for use in space travel, we will be able to provide only general guidelines and recommendations. However, these general guidelines still contain valuable information because they help design the framework for the evaluation of specific systems. Two examples are given to illustrate this point. First, we do not know whether a reduction of cabin noise levels from 45 to 40 dBA will make a difference to crew performance. However, we can say that if noise has a negative effect on performance and task management behaviour, it is more likely to manifest itself in secondary tasks under peak demands or in information sampling behaviour than in primary performance or control action. This is important information since it determines the kind of data one needs to collect during system evaluation. Second, we do not know what kind of in-flight training is needed to provide effective maintenance of crew skills. However, we can say that training design needs to give higher priority to diagnostic skills compared to system control skills.

Being aware of the problems associated with the generalisability of findings, in the next section it is attempted to give some practical recommendations for work design in extended spaceflight. If the empirical findings from our experiments are related to appropriate theory and previous research, the value of the practical implications will gain considerably in substance.

8.5 Practical implications for work design in extended spaceflight

In order to give useful practical recommendations, one needs to consider the characteristics of the work environment of a space crew. This is because space differs in many ways from analogue work environments, on which a great deal of the research is based. These differences have important implications for system design.

8.5.1 Level of system control

First, the operators of space systems tend to have better qualifications and higher abilities than the average process control operator. There may be a relationship between skill / ability level and the desired level of control. One may suspect that space crews are probably in need of a higher level of system control than usually conceded to, and found acceptable by, the average process control operator. However, one needs to remember that 'the more control, the better' is not a panacea, and many instances have shown that too much control can have disadvantages (e.g. Reason 1990; Wickens 1992; see also Experiment 1). Second, in space the allocation of control is a three-dimensional issue, rather than two-dimensional as in most terrestrial contexts, since in addition to crew member and machine, ground control may also demand to exert primary control over certain technical systems. This complicates the already difficult issue of functional allocation (see, for example, Kantowitz and Sorkin 1987; see also Chapter 2) since the decision of whether to allocate the human functions to the cabin or to the ground crew (or flexibly between them) has important implications. During past missions inadequate allocation of function has already led to tension between the space crew and ground control (Connors et al 1985). The degree of independence of the space crew from mission control is also expected to increase with augmenting mission length and, notably, with increasing distance from earth. This is analogous to the situation in the military where the commander on the battle field is better informed about the current situation on the ground than the supreme commander at headquarters; hence is likely to take better operational decisions. The same principle should also be reflected in the control and decision-making structure of a spaceflight. Restricting control unnecessarily may lead to unauthorised attempts to regain it. There was evidence of this in our experimental work in the form of delayed acknowledgement of alarms to increase control. Although this represented only a minor violation of work rules without any consequences for overall system performance, it highlighted the risk of more serious violations of work rules (see Hockey and Maule 1995).

8.5.2 System reliability and repair

A second major difference between space and its analogue environments refers to system maintenance and repair. In order to better understand the problem, we need to examine first the concept of system reliability. If we ignore the human operator for a moment and only look at technical system reliability, we are faced with a dilemma. Technically more advanced systems may not necessarily lead to higher reliability for a number of reasons. First, the use of additional components to increase system reliability may have the opposite effect since each new component also carries the risk of failing, and hence overall system reliability may be reduced (Johannsen 1993; Sanders and McCormick 1992). Second, more advanced systems often represent the latest developments, which contain a number of faults that will only become apparent after an extensive period of testing or, indeed, only during use in the target environment under operational conditions. Third, the use of the technically most advanced systems also tend to require a larger availability of spare parts to replace or repair failed components. This would result in an increased mission payload. More complex systems may also have a higher energy consumption, which would also have implications for mission weight. For the design decisions, one also needs to consider the type of mission (see also section 8.6.2). For a Moon mission or an orbital stay it may be possible to supply a spare part within a reasonable period of time (a few weeks), whereas this would not be feasible for an interplanetary flight. Considering all these points, one seriously needs to ask whether a 'medium tech' approach to system design may be preferable to a 'high tech' approach. This is not to advocate the use of 'dated' technology but, more than in system design for terrestrial use, one should examine whether the potential advantages gained through the use of a newly developed system outweighs its potential risks for reliability and its implications for mission payload.

Closely connected to the reliability problem with similar design implications is a further difference between space and its analogue environments. Unlike in terrestrial process control, in spaceflight no specialised technical support is available on site to deal with more difficult disturbances and system failures. Although astronauts can rely on support from ground control, they need to be considerably more self-reliant to deal with problems of this kind. This requires a much better understanding of space systems than one would expect from the average process control operator, who can rely on system engineers and other experts to arrive on site within a very short time. Again, a 'medium tech' approach may be preferable to facilitate maintenance and diagnostic activities.

If one feels very uncomfortable with the idea of not taking advantage of the latest technological developments, one solution may be to distinguish between safety-critical and non-safety critical systems for the design of space systems. Safety-critical systems (e.g. life support system, emergency escape vehicle) require more redundancy (e.g. components are to be arranged in parallel rather than in series) to be built into the system and should be based on the design philosophy outlined above. The non-safety critical systems (e.g. robot to capture satellites, advanced payload management facility) may be designed according to a design philosophy, which takes advantage of the latest technical developments. This two-tier approach would fulfill the strict criteria in safety critical areas while being able to benefit from the potential of the latest technology in the non-critical areas.

8.5.3 Noise

Despite the significant reductions in noise levels in spacecraft that has been achieved over the years (Connors et al 1985), exposure to noise may still have detrimental effects on performance (though in a very subtle form) as a result of the long exposure times. Many measures have been used to reduce noise levels in a spacecraft (sound-dampening of interior, sound-proofing, ear protection, etc.). Despite the success in reducing noise levels overall, one faces the problem that the number of systems in a spacecraft continues to increase, which augments the number of potential sources of noise. Noise levels in spaceflight are currently largely evaluated with regard to sound pressure, which essentially ignores the degree of annoyance associated with the noise source. Perhaps a shift in emphasis is required which moves away from the physical measurement of noise towards an assessment of the crew's perception of noise. Although sound levels in modern spacecraft are low compared to earlier flights, they are still perceived as disturbing (Kass 1990). The annoyance quality of noise is influenced by a large number of acoustic and non-acoustic factors, which are, among others, frequency, risetime, fluctuations in sound level, attitude towards source of noise, time of day, past experience with noise and activity of listener (Sperry 1978). The following strategies are proposed to achieve a reduction of the negative impact of noise, focusing on the level of annoyance.

First, the crew needs to be provided with control over noise sources. This should preferably be active control (i.e. the equipment can be switched off), or if this is not possible (e.g. life support system), some form of passive control (e.g. withdrawal) is necessary. For this purpose, one may divide the working quarters into zones with different noise levels. For example, a low-noise zone can be created, which allows for the completion of mentally demanding tasks (e.g. report writing) that are more easily impaired by noise. This would also have the advantage that a zone with higher noise levels would be available for task for which some background noise appears to be beneficial. Second, if active or passive control over the noise source cannot be achieved, some other changes to the annoyance-inducing factors may be possible (e.g. change of sound frequency or sound level fluctuation). This may improve the perception of the noise source without having to achieve a reduction in actual sound pressure levels.

Considering the empirical findings of this thesis and previous research, one can be reasonably optimistic that the negative effects of noise exposure on long missions can be maintained within acceptable limits if appropriate measures, such as those suggested earlier, are taken. Even if individuals need to carry out a task in a zone with a non-optimal noise level (it is obvious that not all tasks can be carried out flexibly in any zone since they may be dependent on stationary equipment), their demonstrated ability to use adaptive strategies to cope with the adverse impact of noise for a limited period of time, will probably ensure that no critical performance deterioration will occur.

8.5.4 Sleep deprivation

Loss of sleep is a common occurrence during spaceflight and it is extremely unlikely that its antecedents (vibration, etc.; see Chapter 5) can be sufficiently controlled within the near future. Therefore, it would not be sensible to aim for an elimination of the factors impinging on sleep quality. Instead, one should attempt to accommodate the impact of sleep loss on work behaviour by minimising its negative effects. This is analogous to an approach suggested by Frese, Brodbeck, Heinbokel and Moser (1991) in the context of errors, which stresses the reduction of the negative consequences of errors rather than the elimination of errors per se.

As our experimental work has shown, performance can be maintained under sleep deprivation and a more flexible interface appeared to be of some benefit. Sleep deprivation results in changes in control activity and information sampling behaviour because the operator uses those system management strategies that are more suitable to the altered operational state. This requires the incorporation of flexibility into the work design, not only concerning system control but also regarding the work schedule. Moving away from tight and inflexible mission schedules towards a 'shopping list' system has already been proposed (Kass 1990). Generally, one would assume that the crew is in a better position than ground control to judge when is the best time to complete certain tasks. In addition to having discretion over the timing of action, the crew should also have some discretion over the amount of allocated time for a certain task. This would have the advantage of the crew being able to slow down task completion to maintain accuracy of performance, for example, during elevated fatigue levels. In summary, there should not only be increased flexibility with regard to the order of task completion but also with regard to the time allocation per task.

8.5.5 Training

The findings from this thesis also suggest that in order to evaluate a particular training approach, one also needs to assess its effects on long-term skill retention. Furthermore, if the intention is to train operators to use certain control or information sampling strategies, it is advisable to repeat units of this training approach to ensure that these strategies are maintained because they do not appear to be temporally stable. Concerning the differences found between system- and procedure-based training, one may draw the conclusion that the system-based approach is preferable because in more instances it seems to be superior to the procedure-based approach. However, it is suggested here that one may be able to take advantage of the strength of each. Since in most cases there is more than one crew member who is able to carry out a given task, one may use different training approaches for different crew members to gain advantage of the strength of each method. This idea is based on the concept of cognitive diversity (Westerman, Shryane, Crawshaw and Hockey 1997). Research into fault identification in safety-critical software revealed that the kind of system errors discovered by individuals during fault analysis was partly a function of the structure of their mental model of the system (Westerman, Shryane, Crawshaw, Hockey and Wyatt-Millington 1995). Applied to the present case, the advantage of this approach would be that the mental models of individual crew members would differ to a sufficient extent so that they can complement one another's knowledge and strategies if one approach does not lead to the desired outcome. The idea of cognitive diversity is similar to that of group diversity, which has been proposed to enhance overall group effectiveness (Belbin 1981). The idea is that the group will benefit from people with a different background and experience because they are more likely to complement one another and hence may help prevent group think (Janis 1982). However, in contrast to group diversity, the idea proposed here entails the engineering of diversity rather than the selection of diverse individuals, who have already formed knowledge structures. This represents a more active approach in creating knowledge structures that are different from those of other group members, with the goal of complementing one another's knowledge and hence, leading to higher task effectiveness.

The implementation of this idea, however, would considerably change the structure of training for astronauts. Current training approaches consist of teaching the same material to all crew members, with the specialised crew member for a particular task (e.g. payload specialist, pilot) receiving additional training. Essentially, all crew members receive the same kind of training but the specialists receive more of the same compared to the other crew members. It is argued here that each of them should receive a qualitatively different kind of training. This approach would result in increased total group knowledge and would increase the chance that a crew member will be able to solve a problem that cannot be solved by the others.

The proposed approach also entails some drawbacks. It would undoubtedly be more expensive since crew members would need to be trained in smaller groups, or even individually, and therefore a new structure of flight preparation would become necessary. Furthermore, disagreement about the best way of doing things would increase during routine operations. This may result in more friction within the crew but the appreciation and acceptance of the 'elevated knowledge diversity' may increase if its benefits become apparent in off-nominal situations. It would also require fundamental changes to the general structure of training for spaceflight, which has a very strong emphasis on procedures (Connors et al 1985; Kass 1990). To begin with, one could implement the suggested approach in a small number of areas to gain some experience of its operational effectiveness.

8.6 Future directions in human factors space research

With the duration of space missions rapidly increasing, the importance of factors associated with extended involvement has increased. However, since only little systematic psychological research has been carried out in extended spaceflight, the areas that need researching are numerous and vast. This makes it perhaps more difficult to identify the areas that most urgently require attention to solve some of the human factors problems faced. The areas proposed are all considered to be of crucial importance for the future of space travel. These areas are very different in their scope and focus, which means that there is some overlap among them. Some of them represent problems exclusive to extended spaceflight, whereas others are of relevance in a number of applications. They are now discussed in turn.

8.6.1 Micro-gravity

Among all the space stressors, micro-gravity has had a special position. This has partly been due to its strong impact on many pertinent outcome measures in spaceflight. Moreover, micro-gravity is absent in terrestrial work environments and cannot be adequately simulated on earth, hence making it much more difficult to understand its effects. In spaceflight, micro-gravity may affect work performance in a direct and an indirect way. Neither of these has been systematically researched. In its direct form, micro-gravity has an immediate effect on cognitive functioning. In its indirect form, it affects mediating variables (such as well-being), which may have a negative effect on motivation, resulting in impaired performance. There is some evidence for the first type of effect (Manzey et al 1995) as well as for the second (Kanas 1985; Connors et al 1985). Although the symptoms of both may be very similar, each requires a different solution. Performance decrements, which are a result of motivational problems, require measures to increase levels of motivation (habitat changes, more discretion at work, etc.). Measures to combat impaired cognitive functioning need to focus on task-related aspects to reduce cognitive load (e.g. reduction in task difficulty, memory aids, decision support systems).

Although little systematic research has been done so far on these issues, a few studies have shown the way forward in using the space station as a laboratory for research on human skill maintenance (Manzey et al 1995; Benke et al 1993). However, little is known about the interaction of micro-gravity with other space stressors, such as noise, sleep deprivation, confinement and social isolation. As we know from stress research, the interactions of different stressors are often difficult to predict with some showing an aggravation of their effects while others show a mitigation (Hockey 1986). It is acknowledged that carrying out this kind of research in space is extremely expensive but if we wish to advance knowledge in that area and would like to improve the success of future space missions, we cannot avoid to research the relevant issues in their reference environment. Opportunities for this will become more readily available in the future, in particular, with a permanent human presence on the international space station *alpha*, which will provide an ideal testing environment. Earth-based crews can be used as control groups to isolate the effects of micro-gravity.

Current gravity research centres around problems of transition between 0-g and 1-g. However, during a mission to Mars, at least eight different gravity level transitions have to be made. This has obvious implications for hardware design since the natural body posture changes according to gravity levels and current systems are only designed for 0-g and 1-g levels. Although the impact of gravity level changes will probably be less dramatic than during a change from 0-g to 1-g, or reverse, little is known about the impact of an increased frequency of gravity changes during a mission.

8.6.2 Differences between types of mission

In this thesis extended spaceflight was treated uniformly, ignoring the difference between different types of extended missions. This was done because the commonalities between different mission types were considered stronger than the differences between them. Furthermore, since the commonalities have not yet been well researched, it seemed to be a more sensible approach to concentrate efforts on these first, before attempting to improve our understanding of issues that are only relevant to certain types of mission.

The most important distinction to be made is between interplanetary flights and extended stays on space stations (e.g. in orbit or on a natural satellite). These two types of mission differ with regard to distance from Earth, which has important ramifications for two factors: *rescuability* and *self-sufficiency*. Rescuability refers to the possibility of rescuing the entire crew from a dangerous situation. Whereas it might be possible to rescue the crew within a reasonable period of time from an orbital space station, this possibility becomes rather remote during an interplanetary mission. Individual perception of rescuability is likely to affect the crew state during the mission, in particular, during periods of increased threat to the lives of the crew. Closely related to this is the possibility of replacing individual crew members, because they are not able to complete the mission (e.g., a serious disease or the inability to cope with overall demands). The second factor, self-sufficiency, also indicates more difficulties with the interplanetary missions than with stays on orbital space stations. Unlike interplanetary flights, for stays on orbital space stations or on the Moon, emergency relief flights or scheduled supply flights can be planned to deliver food and technical equipment. This has also implications for the mission payload. Overall this shows that interplanetary flights are more difficult to plan and carry out. On the positive side, the crew will probably be more highly motivated (at least during the first flights of this kind) because of the exploratory and ground breaking nature of the mission. A further difference between missions concerns the kind and distribution of tasks (manoeuvring, payload, process control, communication etc.), which depends very much on the goals and nature of that mission. Future research into extended spaceflight needs to be aware of these differences and address them.

Although the exploration of space is still in its infancy and largely characterised by scientific exploration with little immediate applicability for industrial exploitation, this is likely to change over the years with industrialisation becoming increasingly important. This may involve mining operations and food production on the Moon or industrial production on artificial satellites (European Space Agency 1994; Cheston and Winter 1980). Industrialisation will have profound implications at several levels. Although industry is already currently involved in space exploration, control of the space programme is firmly in the hands of national space agencies. With industry gaining more influence, a shift from safety centred criteria to productivity centred criteria is very likely. At the moment it is a matter of national pride to send space crews safely into orbit and return them unharmed, and no space agency can afford any casualties without receiving serious criticism (see Challenger disaster). This is likely to change when private enterprises become involved, since they would be eager to see a return on their investment. This is not to say that they would be immune to public criticism, or would care about the crew to a lesser extent, but one may suspect that the emphasis on safety would not be as great. Industrialisation would also involve a shift to more physical work (at least in its early stages), with construction work requiring the completion of a considerable amount of extra-vehicular activities. One may consider these construction workers as the equivalent to North Sea divers and oil drillers. Their work is characterised by extreme physical demands combined with very dangerous working conditions. As a result, the process of industrialisation would lead to increased crew diversity.

8.6.3 International crews and system design

The importance of accommodating for an international crew has been recognised, with research having addressed a number of issues. Concern was expressed about communication problems arising from crew members speaking different languages and how this would affect work effectiveness (Connors et al 1985). In particular, there was concern that crew members, for whom the mission language (English and Russian in most cases) was identical to their mother tongue, may dominate the crew or have a disproportional influence. Furthermore, misunderstandings may arise from messages that may have different meanings to non-native speakers (Cushner and Brislin 1996). Personal space also plays a major role in extended spaceflight. Individuals from different cultures differ very much with regard to the amount of personal space needed, and also with regard to the personal distance to other people they feel comfortable with (Cushner and Brislin 1996). Closely related to this is the need for privacy, which seems to be more pronounced in individualistic societies (e.g. USA, UK) than in collectivistic societies, such as Japan (Hofstede 1980). This has implications for the interior design of the space vessel, in particular, the living quarters. In his seminal research Hofstede (1980, 1991) identified four dimensions that can be used to describe differences between cultures: (1) power distance, (2) collectivism vs individualism, (3) femininity vs masculinity, and (4) uncertainty avoidance. Implications of this research are that misunderstandings and difficult situations frequently arise when people from dissimilar cultures interpret signs in a different way and show different kind of behaviour.

Whereas in the areas mentioned above at least some research has been carried out, the area of cultural influences on human-machine interaction has hitherto been largely neglected. When guidelines are developed for system design, they are largely based on the values of Western cultures. Does power distance also apply to the interaction between humans and machines? Are people from a culture scoring high on power distance more likely to trust the machine? Do people from a culture scoring low on uncertainty avoidance place a stronger emphasis on establishing procedures and on using them in novel situations? These are very important questions in the context of human-machine-interaction. In the absence of research, it is not clear whether there are any fundamental differences between cultures that may affect performance and system management behaviour. In the research process one may even identify concepts that have not received any attention in Western work environments as they were considered unimportant or did simply not exist. Anthropological research has often demonstrated that the simple adaptation of measurement instruments through translation (e.g., questionnaires), or operator models in the present case, is not sufficient because it considers only issues relevant in the society it was developed in (Hofstede 1991). Knowledge gained in this area is not only of relevance to spaceflight, many other work environments would also benefit from it. For example, nuclear power stations that are designed and built in Western countries, with the model of Western operators in mind, may be exported to a Latin American country, where this model may be less appropriate. Similarly, aircraft in civil aviation are built for a global market, with little consideration given to intercultural differences. Research evidence suggests that when teams (such as flight crews) are involved in the interaction with complex systems, cultural background is a strong determinant of the kind of interaction (Merrit and Helmreich 1996). An American crew may interact very differently with the system compared to a Korean crew. However, it is difficult indeed to assess the effect of these different patterns of interaction on performance.

8.6.4 In-flight monitoring

The European Space Agency has intensively discussed the possibility of developing and implementing an integrated in-flight monitoring system, which would monitor performance, physiological variables and the affective state of the crew throughout a flight (Ursin 1992). From the results of our empirical work, it is argued that in addition to performance, work strategies (i.e. system control strategies and, notably, information sampling behaviour) are an important factor in understanding adaptive changes in task management. By using a personal baseline for each crew member, the in-flight monitoring system is expected to provide an early indication of signs of strain or any impending performance decrement. However, a number of general problems are associated with the introduction of such a system.

The first question is who should have access to this information. ESA has recognised that this is a very sensitive issue that one needs to approach carefully to gain full acceptance of all crew members. It has therefore been suggested that the crew member concerned would be alerted first if a deviation from the normal operational state was detected (Ursin 1992). It might be possible that the crew member is able to take remedial measures without any support from others. If this was not possible, a confidant from mission control should be informed to discuss measures with the crew member concerned to solve the problem. Using the principle of a design feature of AMT, one could define three categories for each parameter of crew member state: white, yellow and red. The white zone would correspond to an optimal crew member state. The yellow zone would correspond to a non-optimal state, of which only the crew member concerned would be notified so that he/she can take some corrective action. The red zone would represent a more serious deviation that would require some action from mission control. However, the crew member should receive the information first to have some time to take measures to correct the situation, and mission control would only be informed if the attempt of the crew member had been unsuccessful.

The second question is how one can collect meaningful data, which actually indicate changes in the operational state of crew members. Physiological data are probably least difficult to collect since the measurement procedures usually do not interfere with work activities. However, problems may arise during the analysis of some physiological measures (blood samples, etc.) since this needs to be carried out on-board, which is time consuming and requires special equipment. Affective state data may be collected via short, computerised questionnaires. This has already been tested during simulated space missions (Værnes et al 1993; Hockey and Sauer 1996), though crew acceptance can be a problem if too many questionnaires are to be completed. However, the biggest problem in this context is the accurate assessment of performance and work management strategies. Ideally, one would like to collect performance measures during the completion of real work tasks to maximise operational fidelity, but this is associated with a number of difficulties. It is often a problem to measure performance accurately and reliably in a real work environment because it takes place in non-standardised situations and suitable dependent measures are not always available. Furthermore, crew member activities vary from day to day, which makes comparisons over time difficult. An alternative is the assessment on computer-based tasks, which would allow more tightly controlled testing conditions. However, this would require extra crew time to be made available for performance testing, which would not be of direct benefit to the progress of the mission. Instead, it would mean that additional cognitive resources would have to be spent on the completion of performance tests rather than on the mission tasks. This highlights the dilemma faced between the direct measurement of real work performance (poor control, inconsistency of operational conditions, difficulty in obtaining performance indicators) and the measurement of performance on computer simulations (extra time and effort requirements, poorer operational fidelity, acceptability problems).

8.6.5 Further important issues for research

We will now briefly consider some further issues in space exploration, which will grow in importance as spaceflight develops further. The use of VE (virtual environment) has a great deal of potential since it allows a more reality-like simulation of work environments (see Kalawsky 1993). Although current VE systems largely focus on the visual modality, attempts have been made to include other modalities (touch, kinaesthetic), which would considerably increase the fidelity of the simulation (Crawshaw, Hockey, Sauer and Taylor 1996). VE systems are more expensive to develop than conventional simulation software (such as AMT), in particular if the system is fully immersive. Attempts have already been made to build systems of this kind (e.g. at ESA's research centre ESTEC a fully immersive visual simulation of the Moon surface has been developed). The obvious advantage of VE as a training tool is that it is more flexible than a conventional mock-up of a space system with regard to parameter changes. In addition, during the mission VE may also be used for recreational purposes to alleviate the stress associated with isolation and confinement (e.g. the astronaut can walk on a beach in the South Sea).

As previous research has demonstrated the increased likelihood of failure of skilled performance during emergency situations (Broadbent, Baddeley and Reason 1990; Wickens 1992; Reason 1990), one also needs to address these issues in extended spaceflight. It has been attempted, to some extent, to simulate dangerous system failures in this thesis (in the form of control panel failures, see Chapter 6). Although this provided some indication of human behaviour in emergency situations, the simulation fell short of an essential feature of a real emergency. As in most simulation tasks, the level of anxiety experienced is lower than in real work situations because of the absence of a real threat to life, risk to machinery and equipment, or to the environment at large. Future research needs to simulate more realistically the stress-inducing aspects of emergency situations. Previous research has shown that performance in fearful situations does not always deteriorate, some individuals show even enhanced performance during emergencies (Christensen and Talbot 1986). The direction of the effect largely depends on the kind of task to be performed and the degree of previous experience in fearful situations (Idzikowski and Baddeley 1983). This raises the question of whether training approaches that are appropriate for routine situations will also be suitable for non-routine situations.

This thesis focused in its empirical work on individual performance. However, many activities in the space environment require the coordinated effort of several individuals. Although there is a considerable amount of research on work group effectiveness (e.g., Handy 1985), little systematic work in this area has been carried out in spaceflight. Most group research in space focuses on group interaction and behaviour rather than group performance (see Sandal, Værnes and Ursin 1995). This may be another research area that can be explored within the micro-world paradigm. The AMT environment (or any other suitable simulation software) can be modified for individuals working together on the same task while using different PCs. For example, lab-based PCs may be linked to simulate the interaction between mission control and the cabin crew during distributed task completion. This would also allow the examination of issues surrounding three-way functional allocation while using a generic approach.

8.7 Résumé and conclusions

When comparing the challenges and problems faced on extended space missions with the amount of knowledge we have, it is probably not an exaggeration to describe human factors research into space travel as being in its infancy[4]. Because of considerable technical problems, the international spaceflight community does not envisage sending a crew on a Mars mission before the year 2018 (Andrews 1995), and a permanent base on the Moon will not be established before 2006 (European Space Agency 1994). This time period would allow both human factors and psychology to make a solid contribution to a reduction of the risks associated with missions of this kind. This appears to be a realistic task, even if one takes into account that space research operates on time scales of considerable length.

Current endeavours in human factors space research appear not be very systematic. For example, control groups are often not used for reasons of additional costs. If a more systematic approach was adopted, the first step would be to estimate the effect size of different performance shaping factors (including interactions). This would permit the prioritisation of the research programme. The performance shaping factors could then be investigated by using a vertical research approach, that is, the examination of the same phenomena with different methods along the operational fidelity continuum.

The comparative evaluation of the space environment and some suitable analogues suggested that enormous gains can be made from multi-directional knowledge transfer among work environments. We may also accrue benefits from carrying out more space-related research in the most suitable analogue environments, rather than using those which are characterised by easy access but only limited relevance. Central to the degree of similarity is the level of automation found in the respective work environment. Automation is one of the pivotal features of the space work environment. Therefore, it is of great importance to gain an understanding of how the space-related performance shaping factors interact with issues surrounding human-machine-system performance. Using the micro-world approach for generic research in human factors of spaceflight appeared to have been reasonably successful. It allowed us to simulate the space work environment with a considerable degree of fidelity without sacrificing too much control over the experimental conditions. The task environment developed can be easily enhanced for use in a variety of situations. The experimental work has demonstrated that individuals attempt to maintain primary task performance under difficult working conditions and with sub-optimal operator states; and they manage to do so under most circumstances. Since this is usually achieved at a cost, we need to devise systems that are able to monitor whether these costs are excessive, that is, they may entail after-effects or risks to the individual's health, which threatens performance maintenance in the long run. The problems associated with this are great but we can reasonably hope that with further progress in computer technology, in-flight monitoring will be facilitated, notably for physiological and performance measures.

It is hoped that this PhD thesis has contributed to the advancement of knowledge about human factors of space travel. Looking at the work presented, the contribution appears to be rather modest in the light of the challenges faced in future space missions. On the other hand, it may have provided some suggestions as to how such problems might be addressed in a combination of laboratory work and field experiments. Given suitable cooperation with the various space agencies, it should be possible to solve at least some of these human factors questions.

Endnotes:

---

[1] This analysis is restricted to those work environments that are characterised by high automation and / or high complexity and upon which this thesis focuses (i.e. aviation, process control, air traffic control etc.). In some work environments even a primary performance breakdown would cause only minor disruptions (e.g., where unskilled workers can be quickly replaced and where the potential damage to material and machinery is limited).

[2] A detailed treatment of the advantages and disadvantages of using dual-task environments is given in Wickens (1992) and also covered in Chapters 2 and 3 of this thesis.

[3] One could describe the performance-resource function for each task in a multi-dimensional space (MDS). The scores of all AMT sub-tasks represent the coordinates of a point in the MDS, which would have five dimensions in the present case. Every time a performance measurement is taken, a new point is created in the MDS. Variations in performance can be expressed by vectors linking the points in the MDS. The vectors would express performance changes on any of the sub-tasks. In analysing the vectors, one might be able to observe changes in performance as a result of reallocation of resources or changes in working conditions. However, it is not clear whether the suggested approach would actually work in practice. It would need to be tested first in a simpler task environment than AMT (e.g., a triple-task environment with one primary and two secondary tasks) because the picture is likely to be become increasingly complicated with every additional task.

[4] While completing the final draft of the thesis, a book, *'L'homme dans l'espace'* (Rivolier 1997), was published, focusing on psychosocial problems associated with living and working in space. This is the first book that attempts to provide a broader treatment of psychological issues in space since the publication of *Living aloft* (Connors, Harrison and Akins 1985) 12 years ago. It also discusses some of the simulation studies reviewed in Chapters 1 and 7 of the thesis. Rivolier's review addresses issues far broader than the present work (e.g., including an extensive treatment of selection issues and crew compatibility problems).

References

Allport, D.A., Antonis, B. and Reynolds, P. (1972). On the division of attention: a disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology,* 24, 255-265.

Andrews, A. (1995). The rocket man. *New Scientist,* 1983, 27-30.

Angus, R. G. and Heslegrave, R. J. (1985). Effects of sleep loss on sustained cognitive performance during a command and control simulation. *Behavior Research Methods, Instruments and Computers,* 17, 55-67.

Annett, J. (1989). Trained skilled performance. In A. M. Colley and J. R. Beech (eds): *Acquisition and performance of cognitive skills.* Chichester: John Wiley.

Baddeley, A.D. (1972). Selective attention and performance in dangerous environments. *British Journal of Psychology,* 63, 537-546.

Bailey, R. W. (1989). *Human performance engineering.* London: Prentice Hall.

Bainbridge, L. (1974). Analysis of verbal protocols from a process control task. In E. Edwards and F.P. Lees (eds.): *The Human Operator in Process Control.* London: Taylor and Francis.

Bainbridge, L. (1987). Ironies of Automation. In J. Rasmussen, K. Duncan and J. Leplat (eds.): *New Technology and Human Error.* Chichester: John Wiley.

Baltes, P.B. and Willis, S.L. (1982). Plasticity and enhancement of intellectual functioning in old age: Penn State's Adult Development and Enrichment Program (ADEPT). In F. I. M. Craik and S.E. Trehub (eds): *Aging and cognitive processes.* New York: Plenum Press.

Barlow, D. H. and Hersen, M. (1984). *Single case experimental designs.* New York: Pergamon Press.

Bekey, G. A. (1970). The human operator in control systems. In G. B. DeGreen (ed): *Systems psychology*. New York: McGraw Hill.

Belbin, R.M. (1981). *Management Teams.* Heineman.

Benke, T., Koserenko, O., Watson, N.V. and Gerstenbrand, F. (1993). Space and cognition: the measurement of behavioral functions during a 6-day space mission. *Aviation, Space and Environmental Medicine*, 64, 376-379.

Bergan, T., Sandal, G., Warncke, M., Ursin, H. and Værnes, R. J. (1993). Group functioning and communication. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine*, Vol. 3. Greenwich: JAI Press.

Berry, D.C. and Broadbent, D. E. (1984). On the relationship between task performance and associated verbalisable knowledge. *Quarterly Journal of Experimental Psychology*, 36, 209-231.

Berry, D.C. and Dienes, Z. (1993). *Implicit Learning*. Lawrence Erlbaum Associates.

Billings, C. E. (1991). *Human centred aircraft automation: a concept and guidelines.* NASA Ames TM 103885, Moffet Field.

Blackwell, P. J. and Belt, J. A. (1971). Effect of differential levels of ambient noise on vigilance performance. *Perceptual and Motor Skills*, 32, 734.

Bluth, B. J. and Helppie, M. (1987). *Soviet Space Stations as Analogs*. (NASA Grant NAGW-659). Washington DC: NASA.

Bonting, S. L. (ed.) (1993). *Advances in Space Biology and Medicine*, Vol. 3. Greenwich: JAI Press.

Bonting, S. L. (ed.) (1996). *Advances in Space Biology and Medicine*, Vol. 5. Greenwich: JAI Press.

Bourke, P.A., Duncan, J. and Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *The Quarterly Journal of Experimental Psychology,* 49A (3), 525-545.

Brehmer, B. (1990). Towards a taxonomy of microworlds. In J. Rasmussen, B. Brehmer, M. de Montmollin and J. Leplat (eds): *Taxonomy for analysis of work domains. Proceedings of the first Mohawc Workshop.* Roskilde: Risö National Laboratory.

Brehmer, B. (1992). Dynamic decision-making: human control of complex systems. *Acta Psychologica,* 81, 211-241.

Brehmer, B. and Allard, R. (1991). Dynamic decision making: the effects of task complexity and feedback delay. In J. Rasmussen, B. Brehmer and J. Leplat (eds): *Distributed decision-making: cognitive models for cooperative work.* New York: John Wiley.

Brehmer, B. Leplat, J. and Rasmussen, J. (1991). Use of simulation in the study of complex decision making. In J. Rasmussen, B. Brehmer and J. Leplat (eds): *Distributed decision-making: cognitive models for cooperative work.* New York: John Wiley.

Broadbent, D. E. (1976). Noise and the details of experiments: a reply to Poulton. *Applied Ergonomics,* 7, 231-235.

Broadbent, D. E., Fitzgerald, P. and Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology,* 77, 33-50.

Broadbent, D.E., Baddeley, A. and Reason, J. (eds). (1990) *Human factors in hazardous situations.* Clarendon Press: Oxford.

Campbell, S. S. (1992). Effects of sleep and circadian rhythms on performance. In A. P. Smith and D. M. Jones (eds): *Handbook of human performance, Vol. 3.* London: Academic Press.

Carskadon, M.A and Roth T. (1991). Sleep restriction. In Monk, T (Ed): *Sleep, Sleepiness and Performance.* New York: Wiley.

Casali, J. and Wierwille, W. (1984). On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics,* 27 (10), 1033-1050.

Cheston, T. S. and Winter D.L. (1980). *The human factors in outer space production.* Boulder: Westview Press.

Child, J. (1984). *Organization: a guide to problems and practice.* London: Harper & Row.

Christenson, J. M. and Talbot, J. M. (1986). A review of the psychological aspects of space flight. *Aviation, Space and Environmental Medicine,* 57, 203-212.

Chubb, G. P., Laughery, K. and Pritsker, A. A. B. (1987). Simulating manned systems. In G. Salvendy (ed): *Handbook of Human Factors.* New York: John Wiley.

Claxton, G. (1980). Cognitive psychology: A suitable case for what sort of treatment? In G. Claxton (Ed): *Cognitive Psychology: new directions.* London: Routledge.

Cockburn, J. and Smith, P. T. (1994). Anxiety and errors in prospective memory among elderly people. *British Journal of Psychology,* 85(2), 273-282.

Cohen, G. (1989). *Memory in the real world.* Hove: Lawrence Erlbaum.

Cohen, S. and Spacapan, S. (1978) The after-effects of stress: an attentional interpretation. *Environmental Psychology and Nonverbal Behavior,* 3, 43-56.

Cohen, S., Glass, D.C. and Phillips, S. (1979). Environment and health. In H.E. Freeman, S. Levine and L.G. Reader (eds): *Handbook of Medical Sociology.* London: Prentice Hall.

Cohen, S., Evans, G.W., Krantz, D. S. and Stokols, D. (1980). Physiological, motivational and cognitive effects of aircraft noise on children: moving from the laboratory to the field. *American Psychologist,* 35, 231-243.

Connors, M. M., Harrison, A. A. and Akins, F. R. (1985). *Living aloft: Human requirements for extended spaceflight.* Washington DC: NASA Scientific and Technical Information Branch.

Craik, K. (1943). *The nature of explanation.* Cambridge: University Press.

Crawshaw, C.M., Hockey, G.R.J., Sauer, J. and Taylor, P. (1996). Motor control in a tactile virtual environment. Paper submitted to Applied Ergonomics.

Crossman, E.R.F.W. and Cooke, J.E. (1974). Manual control of slow-response systems. In E. Edwards and F.P. Lees (eds.): *The Human Operator in Process Control.* London: Taylor and Francis.

Cushner, K. and Brislin, R.W. (1996). *Intercultural interactions: a practical guide.* London: Sage.

Daniellou, F. (1986). *L' opérateur, la vanne, l' écran: L' ergonomie des salles de contrôle.* Montrouge: ANACT.

Davis, D.R. (1948). The disorder of skill responsible for accidents. *Quarterly Journal of Experimental Psychology,* 1, 136-142.

Davies, A.D.M. and Davies, D.R. (1975). The effects of noise and time of day upon age differences in performance at two checking tasks. *Ergonomics,* 4, 321-336.

Davies, D.R. and Parasuraman, R. (1982). *The psychology of vigilance.* London: Academic Press.

Deaton, J.E. and Parasuraman, R. (1986). Effects of task demands and age on vigilance and subjective workload. *Proceedings of the 32nd annual meeting of the Human Factors Society* (pp 1458-1462). Santa Monica: Human Factors Society.

De Keyser, V. (1986). Technical assistance to the operator in case of incident: some lines of thought. In E. Hollnagel, G. Mancini. D.D. Woods (eds): *Intelligent decision support systems in process environments.* Berlin: Springer.

De Kleer, J. and Brown, J. S. (1983). Assumptions and ambiguities in mechanistic mental models. In D. Gentner and A.L. Stevens (eds): *Mental models.* Hillsdale: Lawrence Erlbaum.

Dinges, D.F. (1992). Probing the limits of functional capacity: the effects of sleep loss on short-duration tasks. In R.J. Broughton and R.D. Ogilvie (eds): *Sleep, Arousal and Performance.* Boston: Birkhäuser.

Dinges, D.F. and Kribbs, N.B. (1991). Performing while sleepy: effects of experimentally induced sleepiness. In Monk, T (Ed): *Sleep, Sleepiness and Performance.* New York: Wiley.

Dinges, D. F., Whitehouse, W. G., Orne, E. C. and Orne, M. T. (1988). The benefits of a nap during prolonged work and wakefulness. *Work and Stress,* 2, 139-153.

Dobson, A. P. (1994). Knowledge and the control of a complex system. Unpublished MSc thesis, University of Hull.

Dörner, D. (1987). On the difficulties people have in dealing with complexity. In J. Rasmussen, K. Duncan and J. Leplat (eds.): *New Technology and Human Error.* Chichester: John Wiley.

Dörner, D. (1990). The logic of failure. In D.E. Broadbent, A. Baddeley and J. Reason (eds): *Human factors in hazardous situations.* Clarendon Press: Oxford.

Dörner, D. and Pfeifer, E. (1993). Strategic thinking and stress. *Ergonomics*, 36, 1345-1360.

Dörner, D., Kreuzig, H. W., Reither, F. and Stäudel, T. (1994). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität.* Bern: Hans Huber Verlag.

Duncan, K.D. (1971). Long-term retention and transfer of an industrial search skill. *British Journal of Psychology*, 62, 439-448.

Duncan, K.D. (1986). Panel discussion on cognitive engineering. In E. Hollnagel, G. Mancini. D.D. Woods (eds): *Intelligent decision support systems in process environments.* Berlin: Springer.

Efimov, V., Gushin, V. I. and Smirnova, T. (1995). Subject's perception of the crew interaction dynamics under prolonged isolation. Paper presented at the HUBES Symposium, 27-28.11.95, Paris, European Space Agency.

Ellis, J.A. (1988). Memory for future intentions: investigating pulses and steps. In M.M. Gruneberg, P.E. Morris, and R.N. Sykes (eds): *Practical aspects of memory: current research and issues: Vol. 1. Memory in everyday life.* Chichester, UK: John Wiley.

Encyclopaedia Americana (1986). Danbury: Grolier.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the 32nd annual meeting of the Human Factors Society* (pp 97-101). Santa Monica: Human Factors Society.

Endsley, M. R. (1995). Towards a theory of situation awareness in dynamic systems. *Human Factors,* 37 (1), 32-64.

Endsley, M. R. and Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors,* 37 (2), 381-394.

European Space Agency. (1994). Manned Aspects of Lunar Environment. Report of MALE-1 Meeting, Avignon (France), 19-21 Apr 94.

Eysenck, M. W. (1984). *A handbook of cognitive psychology.* London: Lawrence Erlbaum.

Eysenck, M. W. and Keane, M. T. (1990). *Cognitive Psychology.* London: Lawrence Erlbaum.

Fitts, P. M. (1951). (ed) *Human engineering for an effective air navigation and traffic control system.* Washington: NRC.

Fleishman, E.A. and Parker, J.F. (1962). Factors in the retention and relearning of perceptual-motor skill. *Journal of Experimental Psychology,* 64 (3), 215-226.

Frankenhaeuser, M. (1989). A psychobiological framework for research on human stress and coping. In S. L. Sauter, J. J. Hurrell and C. L. Cooper: *Job Control and Worker Health.* Chichester: Wiley and Sons.

Franzoi, S.L. (1996). *Social Psychology.* London: Brown and Benchmark.

Frese, M. (1989a). Theoretical models of control and health. In S. L. Sauter, J. J. Hurrell and C. L. Cooper: *Job Control and Worker Health.* Chichester: Wiley and Sons.

Frese, M. (1989b). Kontrolle und Tätigkeitsspielraum. In S. Greif, H. Holling and N. Nicholson: *Arbeits- und Organisationspsychologie: Internationales Handbuch in Schlüsselbegriffen.* München: Psychologie Verlags Union.

Frese, M. (1996). Action errors and action management. Seminar paper presented at the Dept of Psychology, University of Hull, 8.2.96.

Frese, M. and Zapf, D. (1994). Action as the core of work psychology: a German approach. In H.C. Triandis, M.D. Dunnette and L.M. Hough (eds): *Handbook of Industrial and Organizational Psychology, Vol. 4.* Palo Alto, CA: Consulting Psychologists Press.

Frese, M., Brodbeck, F., Heinbokel, T. and Moser C. (1991). Errors in training computer skills: on the positive function of errors. *Human-Computer-Interaction,* 6, 77-93.

Gaillard, A. W. K and Steyvers, F. J. J. M. (1989). Sleep loss and sustained performance. In A. Coblentz (ed): *Vigilance et performance de l'homme dans les systèmes automatisés.* Dordrecht: Kluwer.

Gardlin, G. R. and Sitterley, T. E. (1972). *Degradation of learned skills: a review and annotated bibliography.* Boeing Technical Report No D180-15080-1. Seattle: Boeing Co.

Goldbeck, R.A., Bernstein, B.B, Hillix, W.A. and Marx, M.H. (1957). Application of the half-split technique to problem-solving tasks. *Journal of Experimental Psychology,* 53, 330-338.

Goldberg, J. H. and O'Rourke, S.A. (1989). Prediction of skill retention and retraining from initial training. *Perceptual-and-Motor-Skills,* 69, 535-546.

Greif, S. (1991). Stress in der Arbeit. Einführung und Grundbegriffe. In S. Greif, E. Bamber & N. Semmer (eds): *Psychologischer Stress am Arbeitsplatz.* Göttingen: Hogrefe.

Gushin, V.I. (1991). Soviet experiments with simulated isolation: The main results. Paper presented at ISEMSI Symposium, Paris, 25/26.11.91.

Gushin, V.I. (1992). Peculiarities of the Psychological Investigations in the Prolonged Spaceflight. Paper presented at SPD-2 Meeting in Villefranche, France, 24.-27.3.92.

Gushin, V.I. (1994). Russian Simulation Studies. Paper presented at SPD-4 Meeting in Avignon, France, 17-19.4.94.

Gushin, V.I., Kholin, S.F. and Ivanovsky, Y.R. (1993). Soviet Psychophysiological investigations of simulated isolation: some results and prospects. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine*, Vol. 3. Greenwich: JAI Press.

Gushin, V.I., Efimov, V.A. and Smirnova, T.M. (1996). Work capability during isolation. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine*, Vol. 5. Greenwich: JAI Press.

Hacker, W. (1986). *Arbeitspsychologie*. Bern: Huber.

Hagman, J. D. and Rose, A. M. (1983). Retention of military tasks: a review. *Human Factors*, 25, 199-213.

Handy, C. B. (1985). *Understanding organizations*. London: Penguin.

Harris, J.E. and Wilkins, A.J. (1982). Remembering to do things: a theoretical framework and an illustrative expe riment. *Human Learning,* 1, 123-136.

Harrison, A. A., Clearwater, Y. A. and McKay, C. P. (eds.). (1990). From Antarctica to Outer Space: Life in Isolation and Confinement. New York: Springer Verlag.

Harss C., Lichtenfeld J., Kastner M. and Goodrich J. (1991). In Wise, J.A., Hopkin, V.D. and Smith, M.L. (eds): Automation and Systems Issues in Air Traffic Control. Heidelberg: Springer.

Hart, S. G. (1975). Time estimation as a secondary task to measure workload. Proceedings of the 11th annual conference on manual control. Washington, DC: US Government Printing Office.

Hartley, L. R., Morrison, D. and Arnold, P. (1989). Stress and skill. In A. M. Colley and J. R. Beech (eds): *Acquisition and performance of cognitive skills*. Chichester: John Wiley.

Hoc, J.-M. and Moulin, L. (1994). Rapidité du processus contrôlé et planification dans un micro-monde dynamique. *L'Année psychologique*, 94, 521-552.

Hockey, G. R. J. (1970a). Effect of loud noise on attentional selectivity. *Quarterly Journal of Experimental Psychology*, 22, 28-36.

Hockey, G. R. J. (1970b). Signal probability and spatial location as possible bases for increased selectivity in noise. *Quarterly Journal of Experimental Psychology*, 22, 37-42.

Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue and circadian rhythms. In Boff, K.R., Kaufman, L. and Thomas, J.P. (eds): *Handbook of Perception and Human Performance.* New York: Wiley.

Hockey, G. R. J. (1993). Cognitive-energetical mechanisms in the management of work demands and psychological health. In A.D. Baddeley and L. Weiskrantz (eds): *Attention, Selection, Awareness and Control: A tribute to Donald Broadbent.* Oxford: University Press.

Hockey, G. R. J. (1996). Skilled performance and mental workload. In P.B. Warr (ed): *Psychology at work.* Harmondsworth: Penguin.

Hockey, G. R. J. and Hamilton, P. (1983). The cognitive patterning of stress states. In G. R. J. Hockey (Ed): *Stress and Fatigue in Human Performance.* Chichester: Wiley.

Hockey, G. R. J. and Wiethoff, M. (1993). Cognitive Fatigue in Complex Decision-making. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine*, Vol. 3. Greenwich: JAI Press.

Hockey, G. R. J. and Maule, A. J. (1995). Unscheduled manual interventions in automated process control. *Ergonomics,* 38, 2504-2524.

Hockey, G. R. J. and Sauer, J. (1996). Cognitive fatigue and complex decision-making under prolonged isolation and confinement. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine*, Vol. 5. Greenwich: JAI Press.

Hockey, G. R. J., Sauer, J. and Wastell, D. (1995). Use of a decision-making task to assess cognitive strain and fatigue. Paper presented at the HUBES Symposium, 27-28.11.95, Paris, European Space Agency.

Hofstede, G. (1980). *Culture's consequences: international differences in work-related values.* Beverly Hills: Sage.

Hofstede, G. (1991). *Culture and organizations: software of the mind.* London: McGraw-Hill.

Holding, D. H. (1983). Fatigue. In G. R. J. Hockey (Ed): *Stress and Fatigue in Human Performance.* Chichester: Wiley.

Horne, J. A. (1991). Dimensions to sleepiness. In Monk, T (Ed): *Sleep, Sleepiness and Performance.* New York: Wiley.

Houston, J. P. (1986). *Fundamentals of learning and memory.* New York: HBJ.

Howell, D. C. (1992). *Statistical methods for psychology.* Boston: PWS-Kent Publishing.

Hunt, S. R. (1987). Human engineering for space. In G. Salvendy (ed): *Handbook of Human Factors*. New York: John Wiley.

Idzikowski, C. and Baddeley, A. (1983). Fear and dangerous environments. In G. R. J. Hockey (Ed): *Stress and Fatigue in Human Performance.* Chichester: Wiley.

Janis, I.L. (1982). *Groupthink.* Boston: Houghton Mifflin.

Jerison, H.J. (1957). Performance on a simple vigilance task in noise and quiet. *Journal of the Acoustical Society of America*, 29, 1163-1165.

Jerison, H.J. (1959). Effects of noise on human performance. *Journal of Applied Psychology,* 43, 96-101.

Johannsen, G. (1993). *Mensch-Maschine-Systeme.* Heidelberg: Springer Verlag

Johnson, S. L. (1981). Effects of training device on retention and transfer of a procedural task. *Human Factors*, 1981, 23, 257-272.

Johnson, W.B. and Rouse, W.B. (1982). Training maintenance technicians for trouble-shooting: Two experiments with computer simulations, *Human Factors,* 24, 271-276.

Johnson-Laird, P.N. (1993). *The computer and the mind.* London: Fontana.

Jones D. M. (1983). Noise. In Hockey, G. R. J (Ed): *Stress and Fatigue in Human Performance.* Chichester: Wiley.

Jonsson, A. and Hansson, L. (1977). Prolonged exposure to a stressful stimulus (noise) as a cause of raised blood pressure in man. *Lancet,* 3, 86-87.

Jordan, N. (1963). Allocation of functions between man and machines in automated systems. *Journal of Applied Psychology,* 47 (3), 161-165.

Kahneman, D. (1973). *Attention and effort.* New Jersey: Prentice Hall.

Kalawsky, R. S. (1993). *The science of virtual reality and virtual environments.* Wokingham: Addison-Wesley.

Kanas, N. (1985). Psychological factors affecting simulated and actual space missions. *Aviation, Space and Environmental Medicine*, 56, 806-811.

Kanas, N. (1987). Psychological and interpersonal issues in space. *The American Journal of Psychiatry,* 144, 703-709.

Kantowitz, B.H.(1987). Mental workload. In Hancock, P.A. (ed): *Human Factors Psychology*. Amsterdam: Elsevier Science.

Kantowitz, B.H. and Sorkin, R. D. (1987). Allocation of functions. In G. Salvendy (ed): *Handbook of Human Factors*. New York: John Wiley.

Karasek, R. A. (1979). Job demands, job decision latitude and mental strain: implications for job redesign. *Administrative Science Quarterly,* 24, 285-308.

Kass, J. R. (1990). Study of Columbus crew on-board tasks. Contract report for ESTEC, No. 8548/89/NL/IW, Noordwijk.

Kelly, A.D. and Kanas, N. (1992). Crew member communication in space: a survey of astronauts and cosmonauts. *Aviation, Space and Environmental Medicine,* 63 (8), 721-726.

Kerstholt, J.H., Passenier, P.O., Houttuin, K. and Schuffel, H. (1996). The effect of a priori probability and complexity on decision making in supervisory control task. *Human Factors,* 38 (1), 65-78.

Kessel, C.J. and Wickens, C.D. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. *Human Factors*, 24, 49-60.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: toward an integrative approach. *Psychological Bulletin,* 107 (3), 296-310.

Landeweerd, J.A (1979). Internal representations of a process fault diagnosis and fault correction. *Ergonomics*, 22, 1325-1336.

Lavitola M.S., Tomatis C., Loria A., and Pinotti R. (1990). Terrestrial Analogs as a basis for planning new space habitats. Proceedings of Space Habitability Workshop. ESTEC, Noordwijk, Holland, 28-30.3.90.

Lee, J. D. and Moray, N. (1994). Trust, self-confidence and operators' adaptation to automation. *International Journal of Human-Computer Studies,* 40, 153-184.

Lemoine, M.P. and Debernard, S. (1996). Adaptive assistance to allow dynamic task allocation in the air traffic control. In A.F. Özok and G. Salvendy (eds): *Advances in Applied Ergonomics*. Proceedings of the 1st International Conference on Applied Ergonomics in Istanbul, Turkey, 21-24.5.96. USA Publishing: West Lafayette.

Lorenz, B., Lorenz, J. and Manzey, D. (1996). Performance and brain electrical activity during prolonged confinement. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine,* Vol. 5. Greenwich: JAI Press.

Makhrov, I. V. (1995). Research influences of the factors of spaceflight to the functional characteristics of a cosmonaut in a 'man-machine' system with use of experiment 'Extremal'. Paper presented at the HUBES Symposium, 27-28.11.95, Paris, European Space Agency.

Manzey, D., Lorenz, B. Schiewe, A. Finell, G. and Thiele, T. (1995). Dual-task performance in space: results from a single-case study during a short-term space mission. *Human Factors,* 37 (4), 667-681.

Mateer, C. A., Sohlberg, M. and Crinean, J. (1987). Focus on clinical research: Perceptions of memory function in individuals with closed head injury. *Journal of Head-Trauma Rehabilitation,* 2, 74-84.

Maule, A. J. and Hockey, G. R. J. (1996). The effects of mood on risk-taking behaviour. *The Psychologist,* 9 (10), 464-467.

Mecklinger, A., Friederici, A. D. and Güssow, T. (1996). Attention and mental performance in confinement: evidence from cognitive psychophysiology. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine*, Vol. 5. Greenwich: JAI Press.

Meister, D. (1989). *Conceptual aspects of human factors.* Baltimore, MD: Johns Hopkins Press.

Merrit, A. C. and Helmreich, R. L. (1996). Human factors on the flight deck: the influence of national culture. *Journal of cross-cultural psychology*, 27 (1), 5-24.

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

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

Moray, N. (1986). Monitoring behaviour and supervisory control. In Boff, K.R., Kaufman, L. and Thomas, J.P. (eds): *Handbook of Perception and Human Performance.* New York: Wiley.

Moray, N. and Lee, J. (1990). Trust and allocation of function in the control of automatic systems (EPRL-90-05). Urbana: University of Illinois, Engineering Psychology Research Laboratory.

Moray, N. and Rotenberg, I. (1989). Fault management in process control: Eye movements and action. *Ergonomics,* 32 (11), 1319-1342.

Morgan, G. (1986). *Images of organization.* London: Sage.

Morris, P.E. (1992). Prospective memory: remembering to do things. In M.M. Gruneberg, P.E. Morris, and R.N. Sykes (eds): *Practical aspects of memory: current research and issues: Vol. 1. Memory in everyday life.* Chichester, UK: John Wiley.

Morris, N. M. and Rouse, W.B. (1985). Review and evaluation of empirical research in troubleshooting. *Human Factors,* 27 (5), 503-530.

Morris, N. M., Rouse, W. B. and Fath, J. L. (1985). PLANT: An experimental task for the study of human problem solving in process control. *IEEE Transactions on system, man and cybernetics,* Vol. SMC-15, No 6.

Muir, B. M. (1988). Trust between humans and machines, and the design of decision aids. In E. Hollnagel, G. Mancini and D. D. Woods (eds): *Cognitive engineering in complex dynamic worlds.* London: Academic Press.

Mullaney, D.J., Kripke, D.F., Fleck, P.A. and Johnson, L.C. (1983). Sleep loss and nap effects on sustained continuous performance. *Psychophysiology,* 20, 643-651.

Mundell, I. (1993). Stop the rocket, I want to get off. *New Scientist,* 17.4.93.

Murray, R. L. (1993). *Nuclear energy.* Oxford: Pergamon Press.

Mynatt, C. R., Doherty, M. E. and Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology,* 29, 95-95.

Nagel, D.C. (1987). The Potential of Human Error in Space Operations. Space Life Sciences Symposium: Three Decades of Life Science Research in Space. Washington DC: June 21-26.

Nagel, D. C. (1988). Human error in aviation operations. In E Wiener & D Nagel (eds): *Human factors in aviation.* New York: Academic Press.

National Research Council (1990). *Crew size and maritime safety.* Washington: National Academic Press.

Navon, D. and Gopher, D. (1979). On the economy of the human processing system. *Psychological Review,* 86, 254-255.

Nechaev, A. P., Panfilova, N.G., Sycora, J. and Sholtzova, I. (1995). Investigation of operators' psycho-physiological reserves with the adaptive tracking test. Paper presented at the HUBES Symposium, 27-28.11.95, Paris, European Space Agency.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner and A.L. Stevens (eds): *Mental models*. Hillsdale: Lawrence Erlbaum.

Norman, D. A. (1988). *The psychology of everyday things.* New York: Basic Books.

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

O'Donnell, R. D. and Eggemeier, F. T. (1986). Workload assessment methodology. In Boff, K.R., Kaufman, L. and Thomas, J.P. (eds): *Handbook of Perception and Human Performance.* New York: Wiley.

Ogden, G. D., Levine, J. M. and Eisner, E. J. (1979). Measurement of workload by secondary tasks. *Human Factors,* 21, 529-548.

Oliver, D.C. (1990). Psychological effects of isolation and confinement of a winter-over group at McMurdo Station, Antarctica. In Harrison, A.A., Clearwater, Y.A. and McKay, C.P. (eds.). *From Antarctica to Outer Space: Life in Isolation and Confinement.* New York: Springer Verlag.

Parasuraman, R. (1986). Vigilance, monitoring and search. In Boff, K.R., Kaufman, L. and Thomas, J.P. (eds): *Handbook of Perception and Human Performance.* New York: Wiley.

Parkes, A.M. and Coleman, N. (1990). Route guidance systems: a comparison of methods of presenting directional information to the driver. In E.J. Lovesey (ed): *Contemporary ergonomics 1990.* London: Taylor and Francis.

Pashler, H. (1990). Do response modality effects support multiprocessor models of divided attention? *Journal of Experimental Psychology: Human Perception and Performance,* 18 (4), 826-842.

Patrick. J. (1992). *Training: research and practice.* London: Academic Press.

Patrick, J. and Haines, B. (1988). Training and transfer of fault-finding skill. *Ergonomics,* 31, 193-210.

Pivik, R.T. (1991). The several qualities of sleepiness: psychophysiological considerations. In Monk, T (Ed): *Sleep, Sleepiness and Performance.* New York: Wiley.

Posner, M. I. (1986). *Chronometric explorations of mind.* Oxford: University Press.

Posner, M. I. and Boies, S. J. (1971). Components of attention. *Psychological Review,* 78, 391-408.

Proctor, R. W. and Dutta, A. (1995). *Skill acquisition and human performance.* London: Sage.

Puttkamer von, J. (1988). Der Mensch im Weltraum. In W. Hallmann and W. Ley (eds): *Handbuch der Raumfahrttechnik.* Munich: Hanser Verlag.

Raaijmakers, J.G.W. and Verduyn, W.W. (1996). Individual differences and the effects of an information aid in performance of a fault diagnosis task. *Ergonomics,* 39 (7), 966-979.

Rasmussen, J. (1986). *Information processing and human-machine interaction.* Amsterdam: North-Holland.

Reason, J. (1988). Framework models of human performance and error. In L. P. Goodstein, H. B. Anderson and S. E. Olsen (eds): *Tasks, errors and mental models.* London: Taylor & Francis.

Reason, J. (1990). *Human error.* Cambridge: University Press.

Reid, G. B. and Nygren, T. E. (1988). The subjective workload assessment technique: a scaling procedure for measuring mental workload. In R. A. Hancock and N. Meshkatie (eds): *Human Mental Workload.* Amsterdam: North-Holland.

Rivolier, J. (1997). *L'homme dans l'espace: une approche psycho-écologique des vols habités.* Paris: PUF.

Rivolier, J., Cazes, G. and McCormick, I. (1990). The international biomedical expedition to the Antarctic: psychological evaluations of the field party. In Harrison, A.A., Clearwater, Y.A. and McKay, C.P. (eds.). *From Antarctica to Outer Space: Life in Isolation and Confinement.* New York: Springer Verlag.

Rizzolatti, G. and Peru, A. (1993). Attention during isolation and confinement. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine*, Vol. 3. Greenwich: JAI Press.

Rolfe, J.M. (1973). The secondary task as a measure of mental load. In W.T. Singleton, J.G. Fox, and D. Whitfield (eds): *Measurement of man at work.* London: Taylor and Francis.

Rose A.M. (1989). Acquisition and retention of skills. In G.R. McMillan, D Beevis, E. Salas, M H Strub, R. Sutton and L VanBreda (eds): *Applications of human performance models to system design.* New York: Plenum Press.

Roth, E.M. and Woods, D.D. (1988). Aiding human performance, I: Cognitive Analysis. *Le travail humain*, 51, 39-64.

Rouse, W. B. (1988). Adaptive aiding for human-computer control. *Human Factors*, 30, 431-443.

Sader, M. (1966). *Lautheit und Lärm.* Göttingen: Hogrefe.

Salnitsky, V. P., Dudukin, A.V. and Shlikov, U.V. (1995). Influence of extended subject isolation on the quality of professional activity in the HUBES experiment. Paper presented at the HUBES Symposium, 27-28.11.95, Paris, European Space Agency.

Sandal, G.M. and Berge, B. (1995). Compatibility and interaction among crew members. Paper presented at the HUBES Symposium, 27-28.11.95, Paris, European Space Agency.

Sandal, G.M., Værnes, R.J. and Ursin, H. (1995). Interpersonal relations during simulated space missions. *Aviation, Space and Environmental Medicine,* 66 (7), 617-624.

Sanders, M.S. and McCormick, E. J. (1992). *Human factors in engineering and design.* New York: McGraw-Hill.

Sarter, N.B. and Woods, D.D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors,* 37 (1), 5-19.

Sauer J., Crawshaw M., Hockey G. R. J. and Wastell D. (1996). Performance maintenance on modern ship's bridges: results from two studies using a computer-simulated task environment. In A.F. Özok and G. Salvendy (eds): *Advances in Applied Ergonomics*. Proceedings of the 1st International Conference on Applied Ergonomics in Istanbul, Turkey, 21-24.5.96. West Lafayette: USA Publishing.

Scerbo, M. W. (1994). Implementing adaptive automation in aviation: the pilot-cockpit team. In M. Mouloua & R. Parasuraman (eds): *Human performance in automated systems: current research and trends*. Hillsdale, NJ: Erlbaum.

Sexton, G.A. (1988). Cockpit-crew systems design and integration. In Wiener, E.L. and Nagel, D.C.(eds.): Human Factors in Aviation. London: Academic Press.

Shepherd, A., Marshall, E.C. Turner, A., and Duncan, K. D. (1977). Diagnosis of plant failures from a control plant: a comparison of three training methods. *Ergonomics,* 20, 347-361.

Sheridan, T. B. (1972). On how often the supervisor should sample. *IEEE Transactions on Systems, Sciences, and Cybernetics,* SSC-6, 140-145.

Sheridan, T. B. (1987). Supervisory control. In G. Salvendy (ed): *Handbook of Human Factors*. New York: John Wiley.

Shingledecker, C.A. and Holding, D.H. (1974). Risk and effort measures of fatigue. *Journal of Motor Behavior*, 6, 17-25.

Smith, A. (1989). A review of the effects of noise on human performance. *Scandinavian Journal of Psychology,* 30, 185-206.

Smith, M.J. (1987). Occupational Stress. In G. Salvendy (ed.): *Handbook of Human Factors*. New York: Wiley and Sons.

Spacapan, S. and Oskamp, S. (1990). People's reactions to technology. In S. Oskamp and S. Spacapan (eds.): *People's reactions to technology*. London: Sage.

Spérandio, J.C. (1971). Variation of Operator's Strategies and Regulating Effects on Workload. *Ergonomics*, Vol. 14, No. 5, 571-577.

Spérandio, J.C. (1978). The Regulation of Working Methods as a Function of Workload among Air Traffic Controllers. *Ergonomics*, Vol. 21, No. 3, 195-202.

Spérandio, J.C. (1995). L'analyse de la charge du travail mental. In J.-F. Le Ny and M.-D. Gineste (eds): *La psychologie*. Paris: Larousse.

Sperry, W. (1978). Aircraft and airport noise. In D. Lipscomb and A. Taylor (eds): *Noise control: Handbook of principles and practices*. New York: Van Nostrand Reinhold.

Tattersall, A. J. and Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics,* 39 (5), 740-748.

Tayyari, F. and Smith, J.L. (1987). Effect of music on performance in human-computer-interface. *Proceedings of the 31st meeting of the Human Factors Society.* Santa Monica: Human Factors Society.

Timmermans, D. and Vlek, C. (1992). Multi-attribute decision support and complexity: an evaluation and process analysis of aided versus unaided decision-making. *Acta Psychologica,* 80, 49-65.

Towill, D.R. (1989). Selecting learning curve models for human operator performance. In G.R. McMillan, D Beevis, E. Salas, M H Strub, R. Sutton and L VanBreda (eds): *Applications of human performance models to system design.* New York: Plenum Press.

Tulga, M.K. and Sheridan, T.B. (1980). Dynamic decisions and workload in multi-task supervisory control. *IEEE Transaction on Systems, Man and Cybernetics, SMC-10,* 217-232.

Umbers, I.G. (1979). Models of the process operator. *International Journal of Man-Machine Studies,* 11, 263-284.

Ursin, H. (1992). Integral monitoring in space. In A.W.K. Gaillard (ed): *Integral monitoring in space.* Proceedings of Space Psychology 2, ESA-LTPO, Paris.

Værnes, R.J., Baranov, V.M., Demin, Y.P. and Stepanov, V.A. (1995). HUBES. Report to the European Space Agency, July 1995.

Værnes, R.J., Bergan, T., Lindrup, A., Hammerborg, D. and Warncke M. (1993a). Mental Performance. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine,* Vol. 3. Greenwich: JAI Press.

Værnes, R.J., Bergan, T., Warncke M., Ursin, H., Aakvaag, A. and Hockey, G.R.J. (1993b). Workload and stress: effects on psychosomatic and psychobiological reaction patterns. In S. L. Bonting (ed.): *Advances in Space Biology and Medicine,* Vol. 3. Greenwich: JAI Press.

Wagenaar, W.A. and Groeneweg, J. (1987). Accidents at sea: multiple causes and impossible consequences. *International Journal of Man-Machine-Studies,* 27, 587-598.

Wall, T.D. (1987). New Technology and Job design. In P. Warr (ed.): *Psychology at Work.* London: Penguin.

Wall, T.D and Davids, K. (1992). Shopfloor work organization and advanced manufacturing technology. In C.L. Cooper and I.T. Robertson (eds): *International Review of Industrial and Organizational Psychology. Vol. 7.* Chichester: John Wiley.

Wastell, D. and Sauer, J. (1995). Current and future trends in space psychology: The cognitive ergonomic environment. In G.R.J. Hockey (ed): *Current and future trends in space psychology.* Proceedings of the Space Psychology Days 4, ESA-LTPO Avignon 1994 April 19-21; European Space Agency.

Webb, W.B. and Agnew, H. (1974). The effects of a chronic limitation of sleep length. *Psychophysiology,* 11, 265-274.

Weinstein, N. (1977). Noise and intellectual performance. *Journal of Applied Psychology,* 59 (5), 548-554.

Welford, A.T. (1968). *Fundamentals of skill.* London: Methuen.

Welford, A.T. (1973). Stress and Performance. *Ergonomics,* 16 (5), 567-580.

Welford, A.T. (1980). *Reaction Times.* London: Academic Press.

Welham. M. (1994). *Exploring the deep: The quest to conquer earth's last frontier.* Sparkford: Patrick Stephens Ltd.

Westerman, S.J., Shryane, N.M., Crawshaw, C.M., Hockey, G.R.J. and Wyatt-Millington, C.W. (1995). Cognitive diversity: a structured approach to trapping human error. In G. Rabe (ed): Proceedings of the 14th International Conference on Computer Safety, Reliability and Security, Belgirate, Italy, 11-13 Oct 1995. London: Springer.

Westerman, S.J., Shryane, N.M., Crawshaw, C.M., and Hockey, G.R.J. (1997). Engineering cognitive diversity. Paper presented at the Safety-critical systems symposium '97, Brighton Feb. 1997.

Weybrew, B. B. (1990). Three decades of nuclear submarine research: implications for space and Antarctic research. In Harrison, A.A., Clearwater, Y.A. and McKay, C.P. (eds.). *From Antarctica to Outer Space: Life in Isolation and Confinement.* New York: Springer Verlag.

Weybrew, B. B. and Noddin, E. M. (1979). Psychiatric aspects of adaptation to long submarine missions. *Aviation, Space and Environmental Medicine,* 50 (6), 575-580.

Wichman, H. (1990). Human-machine interaction in spaceflight. In S. Oskamp and S. Spacapan (eds.): *People's reactions to technology.* London: Sage.

Wickens, C. D. (1992). *Engineering psychology and human performance.* Columbus: Merill.

Wickens, C. D., Sandry, D. and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. *Human Factors*, 25, 227-248.

Wiener, E.L. (1987). Fallible Humans and Vulnerable Systems: Lessons learned from Aviation. In J.A. Wise and A. Debons (eds.): *Information Systems: Failure Analysis.* Berlin: Springer.

Wiener, E.L. (1988). Cockpit Automation. In Wiener, E.L. and Nagel, D.C.(eds.): *Human Factors in Aviation.* London: Academic Press.

Wiener, E.L. and Nagel, D.C.(eds.).(1988). *Human Factors in Aviation.* London: Academic Press.

Wientjes, C. J. E. (1992). Respiration in psychophysiology: measurement issues and applications. *Biological Psychology,* 34, 179-203.

Wientjes, C.J.E., Veltman, J.A. and Gaillard, A.W.K. (1995). Cardiovascular and respiratory responses to cognitive task demands during simulation of a 135-day space flight. Paper presented at the HUBES Symposium, 27-28.11.95, Paris, European Space Agency.

Wilkinson, R.T. (1963). Interaction of noise with knowledge of results and sleep deprivation. *Journal of Experimental Psychology,* 66, 332-337.

Wilkinson, R.T. (1992). The measurement of sleepiness. In R.J. Broughton and R.D. Ogilvie (eds): *Sleep, Arousal and Performance.* Boston: Birkhäuser.

Williams, H.L., Lubin, A. and Goodnow, J.J. (1959). Impaired performance with acute sleep loss. *Psychological Monographs: General and Applied,* 73, 1-26.

Wilson, S.L. (1995). Single case experimental designs. In G. M. Breakwell, S. Hammond and C. Fife-Schaw (eds): *Research methods in psychology.* London: Sage.

Winer, B.J. (1970). *Statistical principles in experimental design.* London: McGraw-Hill.

Wirstad, J. (1988). On knowledge structures for process operators. In L.P. Goodstein, H.B. Anderson and S.E. Olsen (eds): *Tasks, errors and mental models.* London: Taylor & Francis.

Wise, J.A., Hopkin, V.D. and Smith, M.L. (eds). (1991). *Automation and Systems Issues in Air Traffic Control.* Heidelberg: Springer.

Woods, D.D. (1986). Paradigms for intelligent decision support. In E. Hollnagel, G. Mancini, D.D. Woods (eds): *Intelligent decision support systems in process environments.* Berlin: Springer.

Woods D., Wise J. and Hanes L. (1981). An evaluation of nuclear power plant safety parameter display systems. In R.C. Sugarman (ed): *Proceedings of the 25th annual meeting of the Human Factors Society.* Santa Monica: Human Factors Society.

Woods D.D., O'Brien, J.F. and Hanes, L.F. (1987). Human Factors Challenges in Process Control: The case of Nuclear Power Plants. In G. Salvendy (ed.): *Handbook of Human Factors*. New York: Wiley and Sons.

Yeh, Y.Y. and Wickens, C.D. (1988). The dissociation of subjective measures of mental workload and performance. *Human Factors,* 30, 111-120.

Yoon, W.C. and Hammer, J.M. (1988). Deep-reasoning fault diagnosis: an aid and a model. *IEEE Transactions on Systems, Man, and Cybernetics*, 18, 659-676.

Appendices

Appendix 1:

AMT 3.0: System faults

- Leak in oxygen valve
- Blocked oxygen valve
- Oxygen valve auto failure (valve permanently open)
- Oxygen set point failure

- Leak in nitrogen valve
- Blocked nitrogen valve
- Nitrogen valve auto failure (valve permanently open)
- Nitrogen set point failure

- Leak in mixer valve
- Blocked mixer valve

- CO_2 srubber ineffective
- CO_2 set point failure
- Vent stuck on

- Cooler set point failure
- Cooler failure

- Dehumidifier set point failure
- Dehumidifier failure

- Control panel failure and nitrogen valve auto failure
- Control panel failure and oxygen valve auto failure
- Control panel failure and cooler failure

Appendix 2:

Fault Finding guide

Symptom : Oxygen levels drop, and may fall below target levels

1.Is there a discrepancy between Oxygen flow meter rates and the rate of fall in Oxygen tank levels?
If *yes* LEAK IN OXYGEN VALVE. *Switch to hi flow and repair.*
If *no* go to 2.

2.Is the Oxygen flow meter reading reduced? (< 6 on default, <9 on med, < 18 on hi)?
If *yes* BLOCK IN OXYGEN VALVE. (The temperature graph will develop plateaux)*.Repair.*
If *no* go to 3.

3. Does the total inflow of Oxygen and Nitrogen to the mixer valve equal the outflow.? If no LEAK IN MIXER VALVE (pressure levels will also be falling)*.Switch both Oxygen and pressure to hi flow and repair)*
.If *yes* go to 4.

4. *Increase Oxygen manually until within target range. Switch to auto. Does* Oxygen level increase above upper limit?
If *yes* OXYGEN SET POINT FAILURE. *Control manually and repair.*

Symptom : Oxygen levels rise above target levels

5. *Decrease Oxygen manually until within target range. Switch to auto.* Does Oxygen level start to rise?
If *yes* OXYGEN VALVE STUCK OPEN. Control *manually and repair*
If *no* go to 6.

6. OXYGEN SET POINT FAILURE. Control *manually and repair.*

Symptom: pressure increase.

7. Does pressure level rise above upper limit?
If *yes* go to 8.
If *no* go to 9.

8. *Switch on vent until pressure is within target range .Switch to auto. Does* pressure continue to fall towards and beyond lower limit?
If *yes* NITROGEN SET POINT FAILURE. Control *manually and repair.*

9. Does pressure graph move in an irregular manner at the upper end of the target range?
If *yes* NITROGEN VALVE STUCK OPEN.(After a while with this fault the temperature graph flattens.) *Repair.*

Symptom: Pressure decreases and may fall below target level.

10. Is the inflow of Oxygen and Nitrogen to the mixer valve greater than the outflow?
If *yes* LEAK IN MIXER VALVE (Oxygen levels will also be falling. ***Switch both Oxygen and pressure to hi flow and repair.***
If ***no*** go to 11.

11. Is pressure falling below target level?
If ***yes increase pressure manually until within target range. Switch to auto.***
Does pressure continue to rise towards and above upper limit
If *yes* NITROGEN SET POINT FAILURE. ***Control manually and repair.***
If ***no*** go to 12

12. Is there a discrepancy between Nitrogen flow meter rates and the rate of fall in Nitrogen tank levels?
If *yes* LEAK IN THE NITROGEN VALVE. Repair.
If ***no*** go to 13

13. Is Nitrogen flow meter reading reduced (<11 on default,< 20 on medium,< 40 on hi)
If *yes* BLOCK IN NITROGEN VALVE. ***Repair***
If ***no*** go to 14

14. Are all flow meter readings 50% below normal?
If ***yes*** go to 15

15. Is Oxygen graph irregular with some drops below target level?
If ***yes*** go to 16

16. Is temperature graph irregular?
If ***yes*** BLOCKED MIXER VALVE. ***Repair***

Symptom: Pressure graph becomes irregular and with reduced amplitude at the lower end of range. Temperature flattens in the top half of the target range.

18. VENT STUCK OPEN. ***Repair***

Symptom: temperature rises above its target state."Cooler on" sign visible.

19. COOLER FAILURE. ***Repair***

Symptom: Carbon Dioxide rises above target levels.

20. CARBON DIOXIDE SCRUBBER FAILURE. Switch *to hi setting and repair.*

Symptom: Temperature falls below target levels."Heater on" sign visible

21.HEATER FAILURE. Repair

Symptom: Humidity rises out of range.

22.*Switch to manual control and bring humidity within range. Switch to auto.*
Does humidity fall below set range?
If *yes* DEHUMIDIFIER SET POINT FAILURE. Control *manually and repair.*
If *no* DEHUMIDIFIER FAILURE. Repair

Symptom: Humidity falls below set range.

23.DEHUMIDIFIER SET POINT FAILURE. Control *manually and repair.*

Appendix 3:

System Knowledge Questionnaire

1.)What functions do each of the following have in the system?

a.) Nitrogen supply

b.) Oxygen supply

c.) Heater

d.) Cooler

e.) Mixer

f.) Vent

g.) Dehumidifier

h.) Carbon Dioxide Scrubber

2.) Under normal conditions the values of the five main system parameters rise and fall in regular patterns. What causes this?

3.) When there is an Oxygen leak the Oxygen first falls below range and then rises slightly and thereafter remains just above the bottom of its range. Can you explain this behaviour?

4.) What states of the system could result in a continuous and steady flow of air through the cabin?

5.) Various types of leak are possible. What exactly happens when there is a leak?

6.) In what way is a cooler failure different from a Carbon Dioxide scrubber failure?

7.) Mark the following statements True "T" or False "F"

a.) The Nitrogen valve opens when pressure falls because the automatic controller responds to the fall in pressure.________

b.) The cooler is automatically activated by the control system when the temperature rises above a set point.__________

c.) The mixer unit operates when activated by the control system.__________

d.) Oxygen flow is reduced when the Nitrogen flow is very high.________

8.) List all the events in the system which could reduce the pressure.

9.) List all the events in the system which could reduce the temperature.

10.) List all the events in the system which could reduce the Carbon Dioxide levels.

11.) List all the system events which could affect the humidity.

Appendix 4:

AMT Questionnaire

J Sauer*
G R J Hockey*
D G Wastell**

*Dept of Psychology, University of Hull
**Dept of Computer Science , University of Manchester

Name:
Date:

Section 1:

The following questions ask about your knowledge of relationships between different processes and system components. Please mark your response in the appropriate box and give a brief explanation.

All questions are presented in a form where arrows are used to describe cause-effect relationships. For example, *Heater (on) ⇒ cabin temperature?* should be read as *What happens to the cabin temperature when the heater is switched on?* The correct response is *Increase* since the operation of the heater causes the temperature to rise.

Example question:
Heater (on) ⇒ cabin temperature ?

Increase	Decrease	Little or no effect
☑	☐	☐

Please explain why: *The heater has a direct effect on the cabin temperature.*

Now to the real questions:

Question 1a:
CO_2 scrubber (on) ⇒ pressure levels ?

Increase	Decrease	Little or no effect
☐	☐	☐

Question 1b:
Please explain why:

Question 2a:
High N_2 inflow ⇒ O_2 inflow ?

Increase	Decrease	Little or no effect
☐	☐	☐

Question 2b:
Please explain why:

Question 3a:
Vent (on) ⇒ pressure levels ?

Increase	Decrease	Little or no effect
☐	☐	☐

Question 3b:
Please explain why:

Question 4a:
Dehumidifier (on) ⇒ O_2 levels ?

Increase	Decrease	Little or no effect
☐	☐	☐

Question 4b:
Please explain why:

Question 5a:
High O_2 inflow ⇒ N_2 inflow ?

Increase	Decrease	Little or no effect
☐	☐	☐

Question 5b:
Please explain why:

Question 6a:
Heater (on) ⇒ humidity levels ?

Increase	Decrease	Little or no effect
[ ]	[ ]	[ ]

Question 6b:
Please explain why:

Question 7a:
Dehumidifier (on) ⇒ temperature ?

Increase	Decrease	Little or no effect
[ ]	[ ]	[ ]

Question 7b:
Please explain why:

Question 8a:
Inflow of N_2 ⇒ temperature ?

Increase	Decrease	Little or no effect
[ ]	[ ]	[ ]

Question 8b:
Please explain why:

Question 9a:
Vent (on) ⇒ N_2 inflow ?

Increase	Decrease	Little or no effect
[ ]	[ ]	[ ]

Question 9b:
Please explain why:

Question 10a:
CO_2 scrubber (on) ⇒ temperature ?

Increase	Decrease	Little or no effect
[ ]	[ ]	[ ]

Question 10b:
Please explain why:

Question 11a:
Inflow of O_2 ⇒ temperature ?

Increase	Decrease	Little or no effect
[ ]	[ ]	[ ]

Question 11:
Please explain why:

Question 12a:
Vent (on) ⇒ O_2 inflow ?

Increase	Decrease	Little or no effect
[ ]	[ ]	[ ]

Question 12b:
Please explain why:

Section 2:

In the following section you are asked in more detail about the relationship between different system components and processes. Please provide as many relationships as you can think of. In order to better understand how to answer the questions, we have given you an example question with answer:

Example Question: Please explain which components or processes have an impact on **humidity** levels, and describe the direction of the relationship.

Example Answer (not necessarily correct):
Dehumidifier: decreases humidity levels
Heater: reduces humidity
N2 inflow: reduces humidity
Cabin crew: increase humidity through breathing

Now to the real questions.

Question 1: Please explain which components or processes have an impact on cabin **pressure** levels, and describe the direction of the relationship.

Question 2: Please explain which components or processes have an impact on cabin **oxygen** levels, and describe the direction of the relationship.

Question 3: Please explain which components or processes have an impact on **temperature** levels, and describe the direction of the relationship.

Thank you for completing this questionnaire.