

Effects of training on short- and long-term skill retention in a complex multiple-task environment

J. SAUER†*, G. R. J. HOCKEY‡ and D. G. WASTELL§

†Institute of Psychology, Darmstadt University of Technology, D-64289 Darmstadt, Germany

‡School of Psychology, University of Leeds, Leeds LS2 9JT, UK

§Department of Computer Science, University of Manchester, Manchester M13 9PL, UK

Keywords: Training; Performance; Skill retention; Multiple-task environment.

The paper reports the results of an experiment on the performance and retention of a complex task. This was a computer-based simulation of the essential elements of a spacecraft's life support system. It allowed the authors to take a range of measures, including primary and secondary task performance, system intervention and information sampling strategies, mental model structure, and subjective operator state. The study compared the effectiveness of two methods of training, based on low level (procedure-based) and high level (system-based) understanding. Twenty-five participants were trained extensively on the task, then given a 1-h testing session. A second testing session was carried out 8 months after the first (with no intervening practice) with 17 of the original participants. While training had little effect on control performance, there were considerable effects on system management strategies, as well as in structure of operator's mental model. In the second testing session, the anticipated general performance decrement did not occur, though for complex faults there was an increase in selectivity towards the primary control task. The relevance of the findings for training and skill retention in real work environments is discussed in the context of a model of compensatory control.

1. Introduction

1.1. *Training and performance*

In many work environments that involve highly complex and automated human-machine systems, there has been a strong interest in the question of what might be the most effective training method to provide operators with the necessary skills. One might argue that the goal of training is generally twofold: it should produce rapid skill acquisition among trainees and also lead to high skill retention during periods of non-practice. Depending on the operational context, one of these two goals may have the stronger emphasis.

For the present purposes, process control provides a good model work environment for the demands placed on the operator during work with complex

*Author for correspondence e-mail: sauer@psychologie.tu-darmstadt.de

and highly automated technical systems. There are three features (e.g. Wickens 1992) that exemplify these distinct skill requirements. First, the process variables controlled have long time constants and therefore respond only very slowly to operator input. Second, it requires the control of a large number of interdependent process variables, which makes it difficult to understand the underlying dynamic (and usually opaque) process. Third, many areas of process control application represent high-risk industries, such as nuclear and chemical plants. The research reported in this article is directly relevant to process control because of the experimental task environment used, which simulates a complex life support system of a space vessel.

Morris and Rouse (1985) identified four training approaches that have been used to teach operators trouble-shooting strategies for the identification and rectification of system faults: (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. (In practice a mixture of these is often used.) 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. Crossman and Cooke 1974, Shepherd *et al.* 1977). This not only applied to familiar faults (i.e. those which were discussed or practised during training), but also to novel faults. However, other work 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). Woods (1988) also expressed concerns about the danger associated with the procedure-based approach during unfamiliar events. However, there seems to be little doubt that extensive practice on the task is a critical factor for the development of a good understanding of how to deal with system disturbances of different kind (De Keyser 1988).

There are also concerns that as a result of automation, traditional training approaches become less effective (Sarter *et al.* 1997). Therefore, operators 'need to form a mental model of the overall functional structure of the system to understand its contingencies and interactions' rather than 'accumulating compartmentalized knowledge' (Sarter *et al.* 1997: 1933). Only an appropriate mental model allows the operator to allocate attention effectively across the different information sources of the system, which is the basis for effective control.

Previous research has often focused on (primary) performance measures to evaluate the effectiveness of training while neglecting other aspects of task management (such as intervention and information sampling strategies) and, indeed, the impact onto the individual's operational state. Furthermore, the retention intervals used during the evaluation process were only of a short duration, which is not always appropriate.

1.2. Skill retention

While the maintenance of acquired skill levels is a desirable goal in any training context, there are some work environments where long-term skill retention is a critical factor. These are characterized by a number of low frequency tasks, which require completion only at certain intervals. The tasks may have to be completed as part of normal operational conditions (shutdown procedure in a processing plant) or they may have to be carried out as part of an emergency procedure (e.g. faultfinding

in nuclear power plant). Examples of work environments with long retention intervals are: the military (Hagman and Rose 1983), extended spaceflight (Sauer *et al.* 1997), shutdown and start-up procedures in process control (Daniellou 1986, Woods *et al.* 1987).

In a recent review of the skill retention literature, seven factors that affect degree of retention were identified (Arthur *et al.* 1998): length of retention interval, task characteristics, instructional strategies or training method, individual differences, degree of overlearning, conditions of retrieval, and testing methods. Particularly relevant to the present work are issues relating to task characteristics and training method. It is not surprising that the type of task characteristics has been found to make a considerable difference. For example, perceptual-motor skills show very little forgetting over long periods, although this appears to be true only of those that require continuous control (Patrick 1992, Swezey and Llaneras 1997). Discrete (procedural) tasks involving step by step actions are more vulnerable to forgetting (Annett 1989, Rose 1989).

Less systematic assessment has been carried out on the effect of training methods, and the literature is generally limited, though it is acknowledged to have an influence on skill retention (Arthur *et al.* 1998). As far as one can tell, however, there have, in any case, been no direct studies of the stability of acquired skills over long periods of retention. One kind of training may be more beneficial in a short-term retention test, and another over a more extended interval. This is especially important in the context of complex tasks, such as process control, where many critical actions need to be carried out only infrequently or during emergencies. In the context of process control, this could mean that operators trained under different regimes (e.g. diagnostic heuristics versus system knowledge) might differ in long-term skill retention (even if acquisition performance and immediate operational effectiveness was identical). Alternatively, any short-term advantage for one kind of training may be lost when skills are retested some months later.

While refresher training is an obvious means to maintain any skill under threat, one still needs to identify the skills most at risk to develop an optimal training programme. In particular, in the area of extended spaceflight, training facilities are limited and training time is at a premium. Furthermore, the problems surrounding skill retention may be aggravated by other factors, such as non-forgiving nature of the tasks (i.e. the first attempt has to succeed) and the non-replaceability of operator.

1.3. *Mental representation of knowledge*

The degree of skill retention and performance in complex physical systems (such as process control) has been associated with the mental model of the operator (e.g. Wickens 1992). The mental model (DeKleer and Brown 1983, Norman 1983) may be considered an intermediate factor in the link between training and performance, with training influencing knowledge representation, which in turn influences performance. 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 (DeKleer 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. The concept of mental model may be a very effective tool because it helps us better understand how training methods influence operator behaviour and performance. In particular in the area of

process control, a good mental model is critical because of the long time constants involved in the process. A good mental model enables the operator to predict and to anticipate the effect of control actions on the system state so that a series of actions can be executed without having to wait for system feedback.

Mental models are multifaceted, with some facets referring to constantly changing knowledge structures (such as current state of the process) whereas others are of a more stable nature. Mental models include different types of knowledge of the physical and functional aspects of the system, such as components, plant layout, cause–effect relationships between parameters, constraints of operations (e.g. maximum allowable levels of parameters) and the current state of the process (Wirstad 1988). It seems plausible to suggest that the stable knowledge structures are primarily determined by the type of training received while knowledge of the current system state is more likely to be a function of system design features such as transparency, feedback and interrogability.

The elicitation of system knowledge from human operators has generally been considered a difficult task (Rouse and Morris 1986). A number of methods have been used in the research literature to elicit knowledge. Rowe *et al.* (1996) identified four general methods: (1) accuracy and time measures—performance measures collected during task completion, (2) structured or unstructured Interviews, (3) process tracing—verbal protocols in which operators think aloud while completing the task and (4) structural analysis of concepts by paired comparison similarity judgements. Naturally, each method is associated with certain advantages and disadvantages (Rowe *et al.* 1996).

For the present study, a combination of the first three of these methods was adopted, based on a distinction between implicit and explicit knowledge (Broadbent *et al.* 1986). Implicit knowledge relates to *accuracy and time measures*, and is derived from performance measures (knowledge that manifests itself in task performance). For the measurement of *explicit knowledge* (knowledge that can be verbalized and explained), the *process tracing* method, supplemented by a variation of the *interview* method, was made use of. The methods *process tracing* and *interview* represent a means to explore underlying cognitive processes in more detail while *accuracy and time measures* offer only indirect evidence for existing knowledge structures (Rowe *et al.* 1996). They are considered complementary methods in gaining access to operator knowledge of a system.

2. The present study

The work aims to compare two fundamentally different training approaches with regard to their effectiveness for short- and long-term skill retention in managing faults of different complexity. Of interest was not only which training method would be best after an extended layoff period, but also the interaction between training method and fault difficulty. A comprehensive methodological approach was employed to evaluate the training methods, which took into account the often subtle adaptation patterns of individuals in response to changes in task demands, which can make them difficult to detect.

The first training approach (called *system-based* training) aimed to provide the operators with a deeper understanding of the relationship and interaction between the components of the system. It was demonstrated how this knowledge was to apply to manage system failures of various kinds. In contrast, the second training approach (called *procedure-based* training) provided the operators with precise procedures to

follow when a system disturbance occurred. This approach belonged to the category 'Guidance in the use of algorithms or rules' in Morris and Rouse's (1985) classification system while the system-based approach corresponded to the category 'Guidance in the use of system knowledge'. Applying Stammers' (1996) model of three levels of knowledge to the present training methods, the procedure-based approach focused more on the lowest knowledge level (*task knowledge*, i.e. how to do something) than the system-based approach. Conversely, the system-based approach placed more emphasis on the higher knowledge levels: *functional knowledge* (why something is done) and *general knowledge* (putting something into a wider context).

2.1. Model of compensatory control mechanisms

The research is embedded within a model of compensatory control (Hockey 1993, 1997). This model argues that humans use various compensatory strategies to maintain performance under stress and high workload. Although decrements in *primary task performance* are sometimes observed, these are rare in operational contexts, probably because of the recruitment of additional effort, particularly where the costs associated with a severe performance breakdown are high. Instead of suffering decrements on primary tasks, individuals may show one or more of a number of *latent decrements*. A common pattern is that of strategic adjustment. Work goals may be lowered to prevent a catastrophic breakdown of performance, or the way in which the task is carried out may be changed to reduce the load on planning and working memory (e.g. the shift by process operators to closed loop control in high workload situations; Bainbridge 1974). In multiple tasks, people may reduce attention to secondary task components, so that decrements are more likely to be found on these. It is also possible to measure the costs of compensatory activity, such as increased ratings of effort and fatigue, or increased levels of sympathetic nervous system activity (e.g. the release of catecholamines). Finally, prolonged periods of sustained effort in the performance protection mode are likely to give rise to fatigue after-effects, which can be detected by the administration of sensitive probe tasks at the end of a work session.

2.2. Multiple-task environment

The model has already been tested in previous studies (e.g. Hockey *et al.* 1998) and proved useful in guiding research in the context of complex task performance. Following from the kind of compensatory control mechanisms predicted by the model, it was necessary to use a task environment that allows these adaptive patterns to occur and to be observed. Based on the characteristics of a process control environment, a computerized simulation task CAMS (Cabin Air Management System) was developed to evaluate these patterns. Since the European Space Agency funded the work, CAMS was modelled on a life support system in a spacecraft (for a detailed description of the CAMS rationale, see Sauer *et al.* 2000).

Real work environments are characterized by a range of tasks that have different priorities attached, that is, some are of greater importance (primary tasks) than others (secondary tasks) for the fulfilment of overall work goals. Therefore, it is important to incorporate this feature into the simulation environment.

This paper reports an experiment that examined the effectiveness of two methods of training, based on low level (procedure-based) and high level (system-based) understanding of complex system. The effectiveness of training is compared across retention intervals of different lengths (short- versus long-term skill retention).

Furthermore, system faults of different difficulty (well-practised, novel and emergency faults) were used to test for differential effects of the training methods.

It was predicted that procedure-based training would lead to better performance for practised faults while system-based training would be beneficial for novel faults and emergencies. This was based on the assumption that procedure-trained operators would be more efficient in the execution of rule-based system management strategies. In contrast, operators trained on the system-based approach were expected to be at an advantage when real-time problem solving was required. It was also predicted that the rule-based execution of system control would require fewer mental resources, which would be reflected in fewer secondary task decrements. Finally, on the basis of the scant previous work on skill retention one would expect procedural training to show generally greater decrements of performance over a long retention interval.

3. Task environment CAMS

3.1. General features

The task environment was an enhanced version of CAMS, which has already been used in studies (Hockey *et al.* 1998, Sauer *et al.* 1999a–c).

The task environment provides a generic simulation of a spacecraft's life support system, running on a (386 or better) IBM-compatible PC. The life support system comprises five subsystems, with the main system variables (O_2 , CO_2 , cabin pressure, temperature, humidity) normally maintained within their target range by automatic controllers. These subsystems are closely coupled and, therefore, have an effect on each other during system operation. While the system is running automatically, the main task of the operator is to monitor the performance of the automatic controllers and to intervene (by adopting manual system control) in the event of a system fault (e.g. ineffective CO_2 scrubber, N_2 leak, O_2 set point failure).

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. This would indicate an elevated risk of a parameter moving into the *red zone*. If a parameter moves into the red zone, it is counted as an error. The boundaries of the red zone were defined: oxygen (19.0–20.5%), pressure (970–1040 mbar), carbon dioxide (0.1–1.5%), temperature (18.5–23.0°C) and humidity (36–44%).

A characteristic of complex systems is that their components do not operate independently from each other. Similarly, CAMS has been designed to incorporate interactions to make the task environment as realistic as possible. The knowledge of these interactions will aid the skilled operator to control the system more efficiently. Because of the limited time available for performance testing, the dynamic processes in the system were greatly accelerated and some interactions were magnified in comparison with the natural laws that represented the basis of the mathematical equations used in the simulation (e.g. Boyles Law: increasing pressure causes a rise in temperature). Each subsystem operates by using set points, which trigger off some regulatory activity once a parameter has reached this prespecified level.

CAMS allows any combination of automatic and manual operation of the system. However, it should be emphasized that full manual control of the system is not a trivial undertaking. The five subsystems are tightly coupled and affect each other's operation. Therefore, if the interdependencies of the subsystems are not well understood, manual control can lead to operator-induced system disturbances. The adoption of manual control attracts some costs in the form of increased attentional resource requirements and a higher risk of errors.

3.2. Operator interface

The main display (figure 1) provides a mimic of the topographical layout of the system, with additional features such as a warning system, history display and system clock. For each parameter a control panel is available, enabling the operator to manipulate the levels by taking over manual control.

The main display also provides 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. The operation of the various subsystems (e.g. CO₂ scrubber, cooler, dehumidifier) is indicated by various icons and messages. The warning system issues an alarm if any of the key parameters moves out of its safe range. This takes the form of a sign being displayed in red colour, indicating the parameter concerned. In addition to genuine alarms, false alarms occurred at a rate of ~0.25/min.

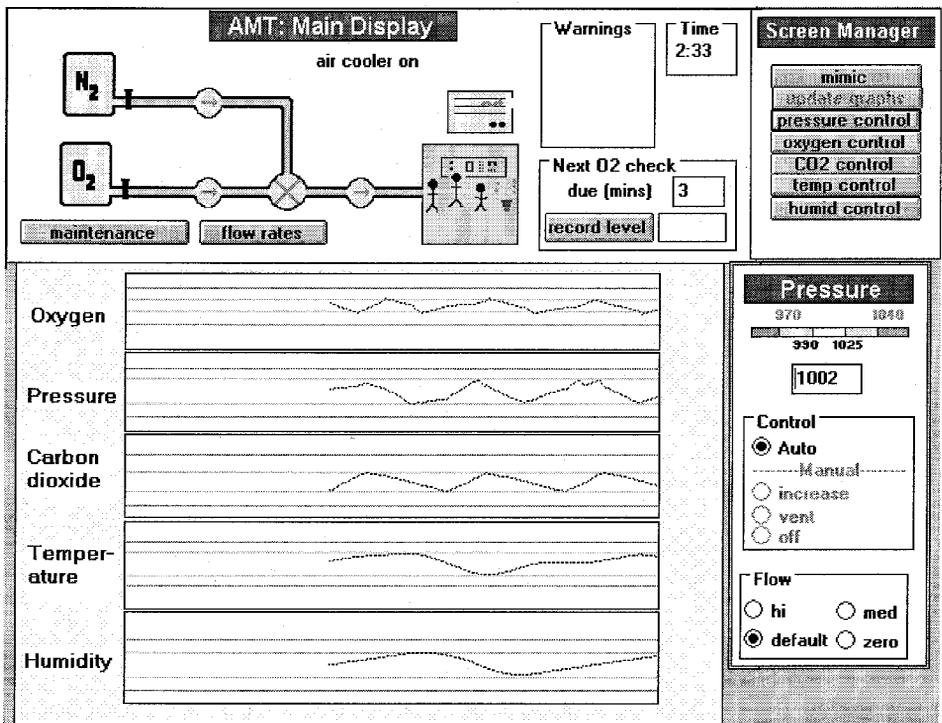


Figure 1. Main display of CAMS.

With the screen manager (figure 1, top right), the operator can invoke different displays (history display, five control panels) by clicking on the appropriate button. The history display (figure 1, bottom left) provides graphical information about the development of each key parameter over the previous 4 min. The maintenance facility (figure 2) is evoked by clicking on the 'maintenance' button, allowing the operator to repair system faults. Each repair takes 1 min to complete during which time no other repair can be carried out. Since no explicit feedback is given whether the diagnosis was correct (as in real work environments), the operator needs to carry out checks to determine whether the fault state has been rectified.

3.2.1. *Tasks of the operator:* Four tasks are to be carried out by the operator, which are classified as primary and secondary tasks according to their importance for crew survival. The first two tasks listed below are defined as primary tasks, whereas the remaining two are considered secondary tasks.

- (1) *System control:* a safety-critical task is to maintain the five key parameters within their target state. This requires the monitoring of safe levels and the completion of interventions if required.
- (2) *Fault diagnosis:* if the operator discovers a system fault, he/she is required to engage in a process of fault diagnosis and repair. To diagnose malfunctions of the system successfully, the operator needs to carry out a number of tests to determine the precise cause of the disturbance.
- (3) *Acknowledgement of alarms:* this task refers to the monitoring of the alarm system and is essentially a reaction time task. If a warning signal appears on

Fault fixing

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.*

close

Figure 2. Maintenance facility of CAMS.

the screen, the operator needs to click upon it immediately. 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.

- (4) *Tank level recordings*: this is essentially a prospective memory task (e.g. Brandimonte *et al.* 1996), where the operator is required to execute an activity at fixed intervals without receiving any further prompt. In the current study, the task was to record O₂ tank levels at 3-min intervals.

3.3. *Special features of CAMS*

3.3.1. *Data-gathering facility*: The data logging facility allows the recreation of the experimental session in its entirety. It automatically records the system state at 10-s intervals. This comprises the level of key parameters and tanks. Every action taken by the operator (e.g. increasing O₂ 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.

3.3.2. *Fault states*: Up to 18 different fault conditions can be programmed into CAMS by the experimenter, specifying the exact time each of them will occur during the testing session. 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 requires 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 faults, control panel failures, refer to a qualitatively different kind of fault. These are much more difficult to manage for two reasons. First, part of the fault state is a complete breakdown of a control panel, which means that the corresponding parameter cannot be controlled any more in a direct way. Second, control panel failures cannot be repaired because of a concurrent failure of the maintenance facility.

3.3.3. *Subjective state measures*: CAMS allows the collection of subjective state measures at fixed intervals during the experimental session by means of visual analogue scales (100-mm lines) that are presented on the computer screen. Participants are asked to indicate their rating of *mental effort*, *anxiety* and *fatigue* by moving the slider along the scale with the mouse. The number of scales was kept small deliberately to minimize disruptions of the task. Previous experience with the use of the scales was positive, with all of them showing a satisfactory scale reliability (Cronbach's $\alpha > 0.85$; Hockey *et al.* 1998).

4. Methods

The work presented here was based on two experimental testing sessions. The first lasted for 3 h, with each 1-h period associated with a different noise condition (quiet, noise, quiet). The effects of noise are published in a separate paper. Here, only the first hour of this testing session (i.e. noise-free) will be examined and the results will be compared with a second testing session, which was completed 8 months after the first to assess the impact of training on long-term skill retention.

4.1. Design

The design employed was a $2 \times 4 \times 2$ mixed factorial design, with training, fault type, and layoff period as independent variables. Type of training was a between-subjects variable, varied at two levels (system-based versus procedure-based). Fault difficulty was varied within sessions at four levels: Fault-free, practised fault, novel fault and control panel failure (the fault types are described in detail below). Layoff period, a within-subjects variable, had two levels: short-term (immediate testing after training) and long-term skill retention (testing after an 8-month layoff period, with no intervening practice).

4.2. Participants

Twenty-five participants (23 male, two female) were recruited for the study, all members of the University of Hull either as final-year undergraduates ($n = 8$), postgraduates (14) or staff (3). Participants were members in (or graduated from) the following departments: Engineering (15), Chemistry (7), Biology (2) and Computer Science (1). All were chosen on the basis of some basic selection principles to ensure that they matched the elementary criteria for the selection of trainee astronauts. Because some of them have left the university during this period, only 17 participants (68%) completed the second testing session. The average age of the participants was 25.1, ranging from 21 to 31 years.

4.3. Training

Participants were randomly assigned to one of the two training regimes. Thirteen participants received *procedure-based training* and 12 were given *system-based training*. Both groups received ~ 6 h of training, which comprised five sessions. The first session was identical for both groups, in which the basic functioning of CAMS was explained. In the following four sessions, both groups were trained on exactly the same system faults but with a different emphasis (see below). Both groups had a faultfinding 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.

4.3.1. Procedure-based training: In this condition the training approach emphasized the importance of following the procedures for fault identification and management, both of which were described in the faultfinding guide. As a result, participants in this group spent more time consulting the guide than the system group. After the participants had become more familiar with the different fault state, the instructor only intervened if they did not follow the prescribed procedures. 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.

4.3.2. System-based training: In this condition participants were encouraged 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. The training sessions were similar to those of the procedure group with regard to the sequence of faults practised. However, one important 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. 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). 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 of the system.

4.4. 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 a concurrent failure of the 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.

4.5. Assessment of mental model

Measuring the operator's knowledge of the system (both implicit and explicit) assessed the operator's mental model. In each case they were tested on their knowledge of the three control panel failures, as these were the most comprehensive fault states used in the study, and those that are most strongly related to dealing with low frequency problems. 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. that the inflow of liquid N₂ will reduce the cabin temperature and can therefore be used in the event of a failed cooling system).

Implicit knowledge was measured by examining performance during the occurrence of the three control panel failures, and by asking participants to give a brief demonstration of their control strategies in the post-experimental interview. Explicit knowledge was measured by means of verbal protocols and a structured questionnaire. The verbal protocol was recorded while the participant managed the system during the three fault states. The verbal protocol was transcribed and then analysed to assess the participant's understanding of system operation. This was supplemented by the results of a questionnaire that measured the operator's understanding of the underlying causes of particular fault states. The questionnaire was in two parts. The first consisted of 12 three-alternative multiple choice items (e.g. What happens to the cabin temperature when nitrogen is pumped into the cabin? (a) Increase, (b) decrease or (c) no effect). Below each item, operators were asked to explain the nature of the relationship in more detail. In the second section, open

questions were asked about three cabin parameters (pressure, oxygen, temperature), in which the operators' mental models were tested by asking them to explain their understanding of the system (e.g. Please explain which components or processes have an impact on cabin oxygen levels, and describe the direction of the relationship).

The evaluation of the mental model was carried out independently by two of the experimenters. The interrater reliability coefficient (Cohen's K) was satisfactory for explicit ($K = 0.88$) as well as for implicit knowledge ($K = 0.83$).

4.6. Procedure

In the experimental session, 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 5 and 10 min, 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 or their duration. The following example illustrates a typical set-up of a 30-min subperiod: fault-free (min 0–3.5), practised fault (min 3.5–11), fault-free (min 11–15), control panel failure (min 15–22), fault-free (min 22–24.5), novel fault (min 24.5–29.5).

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 were presented in the session: block of oxygen valve, O₂ set point failure, N₂ 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), N₂ valve permanently open (with a simultaneous pressure CtrlPanF) and O₂ valve permanently open (with a simultaneous O₂ 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).

The second testing session (after the 8-month layoff) began with a 3-min 'warm-up' session to allow the participants to familiarize themselves again with the task. They were 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, the 1-h testing session was completed, corresponding exactly to the procedure for the first testing session. Following the task session an assessment was made of the mental model of each participant.

5. Results

The data analysis is based on the 17 participants (10 from the procedure group, seven from the system group) that completed both testing sessions. 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 when data are positively skewed (as often occurs when performance is expressed in error terms). However, for reasons of clarity the means presented represent the untransformed values.

5.1. Primary task performance

5.1.1. *Parameter control failures:* This measure refers to the percentage of time each of the five key parameters had been out of range (i.e. in the red zone) during the task sessions. The system group did slightly better than the procedure group (3.7 versus 4.6%), though the difference was not significant ($F = 1.23$; d.f. = 1,15;

$p > 0.05$). This small difference between groups remained stable over the retention interval. It was predicted that ProcG (procedure group) would perform better during PracFs while SysG (system group) would be better at NovFs and CtrlPanFs. While there is some suggestion of this in table 1, the interaction was not significant ($F < 1$). After the layoff period, participants, surprisingly, showed improved performance compared with their first testing session (3.0 versus 5.2%). The main effect of layoff period was significant ($F = 5.51$; d.f. = 1,15; $p < 0.05$). However, as table 1 shows, this was largely due to a strong interaction between fault type and skill retention ($F = 12.5$; d.f. = 3,45; $p < 0.001$). For PracF, there were twice as many errors on the second session (LSD-test $p < 0.05$). No effect was found for novel faults although the data indicate a small improvement. The strongest effect was observed for control panel failures, which were halved (12.9 versus 6.1%; LSD-test $p < 0.001$). As predicted, the data showed a strong main effect of fault type, the number of control failures rising with increased fault difficulty ($F = 96.8$; d.f. = 3,45; $p < 0.001$). *Post-hoc* LSD tests showed that all means were significantly different from each other ($p < 0.05$). Overall, there was no evidence for forgetting of the primary control activities.

5.1.2. *Errors in fault diagnosis*: This measure was scored for the percentage of incorrectly diagnosed faults. Control panel failures were not included in this analysis since they could not be diagnosed as the failure of the maintenance facility was part of the fault state. Most faults were correctly diagnosed at the first attempt in both groups (84.3%). A further 11.3% of faults were correctly diagnosed at a later attempt, leaving 4.4% of faults where a correct diagnosis failed. For this analysis, the strict error criterion (incorrect at first time) was employed, though the less strict criterion did not provide a fundamentally different pattern. ProcG (14.2% incorrect diagnoses) showed a slightly better performance than SysG (17.2%) but the difference was not significant ($F < 1$). Diagnostic performance was also found to have deteriorated slightly after the layoff period (13.1 versus 18.2%) but again, this change was not significant ($F < 1$). As predicted, NovFs were less accurately

Table 1. Effects of training on primary and secondary task performance at T_1 and T_2 (mean score of performance measures).

	System group		Procedure group		T_1 overall	T_2 overall
	T_1	T_2	T_1	T_2		
Percentage of parameter control failures	4.9	2.5	5.6	3.5	5.2	3.0
PracF	0.7	0.9	0.1	0.9	0.4	0.9
NovF	1.5	1.4	3.3	2.4	2.4	1.9
CtrlPanF	12.4	5.2	13.3	7.1	12.8	6.2
Fault identification time (s)	109	113	92	131	101	122
Percentage of omitted tank level recordings	13.8	17.2	8.0	12.7	10.9	14.9
F-free	7.1	7.1	0.0	5.0	3.6	6.1
PracF	14.8	8.7	6.4	20.4	10.6	14.6
NovF	4.8	19.5	15.4	17.9	10.1	18.7
CtrlPanF	28.6	33.3	10.0	7.5	19.3	20.4

F-free, fault free; PracF, practised faults; NovF, novel faults; CtrlPanF, control panel failures.

diagnosed than PracFs (22.5 versus 8.8%). This difference was significant ($F = 7.59$; d.f. = 1,15; $p < 0.05$). No other significant effects were found (all $F < 1$). As with control performance, there was no evidence of forgetting on the diagnostic accuracy measure.

5.1.3. *Fault identification time*: The speed of fault diagnosis was evaluated by measuring the time needed to make a correct diagnosis. Table 1 shows an overall increase in fault identification time at T_2 . This was almost entirely accounted for by ProcG, who took 40 s longer to identify faults on the second session (an increase of 43%). SysG participants showed very stable identification times over the two sessions (109 versus 113 s). While the main effect was significant ($F = 6.35$; d.f. = 1,15; $p < 0.05$), the interesting effect is the significant interaction ($F = 4.83$; d.f. = 1,15; $p < 0.05$).

5.2. Secondary task performance

5.2.1. *Annunciator reaction time*: This measure indicates the time needed (in seconds) to respond to the appearance of an annunciator that warns the operator of the unsafe level of a system parameter. Whereas no main effect of training was found ($F < 1$), the analysis revealed an effect of fault type ($F = 9.13$; d.f. = 3,45; $p < 0.001$). The more difficult the system faults became, the longer participants took to respond to annunciators (F-free: 1.9 s; PracF: 2.0 s; NovF: 2.3 s; CtrlPanF: 2.5 s). *Post-hoc* LSD-tests confirmed that the differences between all four levels of fault type were significantly different from each other (at $p < 0.05$ level or higher). There was, however, no significant effect of layoff period ($F = 3.45$; d.f. = 1,15; $p > 0.05$), RT actually improving between the two sessions (from 2.8 to 2.5 s). No other effects were observed.

5.2.2. *Prospective memory (tank level recording)*: Recordings were assigned to four different categories according to the accuracy of timing: *early* (5 s or more before scheduled time), *on-time* (within 5 s of scheduled time), *late* (between 5 and 20 s after scheduled time) and *omissions* (missed or > 20 s after scheduled time). The occurrence of early responses was generally very low ($< 1\%$). Most recordings were made on time (47%), followed by late responses (39%) and a not inconsiderable number of omissions (13%). The category of omissions emerged as the most sensitive one and was therefore selected for the analysis. The data are presented in table 1. As found for reaction time, fault type had an impact on performance, more recordings being missed during more complex fault states ($F = 3.26$; d.f. = 3,45; $p < 0.05$). The interaction of training and fault type was significant ($F = 3.02$; d.f. = 3,45; $p < 0.05$), though the pattern of results is unexpected. Procedure-trained participants had less omissions during CtrlPanF (8.8%, averaged across T_1 and T_2) compared with the two less demanding fault states (PracF: 13.4%; NovF: 16.7%). This was in contrast to SysG, who followed the expected pattern of increasing omissions with higher fault difficulty (PracF: 11.8%, NovF: 12.2%, CtrlPanF: 31.0%). Again, there were no effects of layoff period ($F = 3.45$; d.f. = 1,15; $p > 0.05$).

5.3. Control actions and information sampling behaviour

5.3.1. *Manual control actions (MCA)*: The data for MCA are presented in table 2. The F-free condition was not included in this analysis since virtually no control actions were carried out during this phase. The analysis of MCA revealed that system-trained participants carried out more than twice as many control actions than

Table 2. Effects of training on control activity and information sampling behaviour at T_1 and T_2 (mean score of intervention and sampling frequency).

	System group		Procedure group		T_1	T_2
	T_1	T_2	T_1	T_2	overall	overall
Manual control actions (interventions/min)	0.59	0.79	0.25	0.78	0.42	0.79
PracF	0.05	0.33	0.06	0.52	0.05	0.42
NovF	0.13	0.14	0.23	0.71	0.18	0.43
CtrlPanF	1.59	1.89	0.47	1.10	1.03	1.50
History display (time on screen in %)	86.0	79.0	88.5	89.0	87.3	84.0
Flow meters (samples/min)	2.41	2.62	1.80	2.19	2.10	2.41
Maintenance facility (samples/min)	0.39	0.24	0.26	0.17	0.32	0.21

PracF, practised faults; NovF, novel faults; CtrlPanF, control panel failures.

their procedure-trained colleagues at T_1 (0.59 versus 0.25/min). This difference disappeared at T_2 where both groups were equally active. The interaction however failed to reach significance ($F = 2.59$; d.f. = 1,15; $p > 0.05$). The effects of fault type are quite complex. There was the expected main effect of fault type, with most MCA being carried out during CtrlPanFs, followed by NovFs and PracFs ($F = 25.22$; d.f. = 2,30; $p < 0.001$). *Post-hoc* LSD-tests revealed that CtrlPanF was significantly different from the two others (both LSD-tests $p < 0.001$). More importantly, the two training groups showed different patterns of interventions with increasing fault difficulty. The increase in MCA for CtrlPanFs was more pronounced for SysG than for ProcG ($F = 7.03$; d.f. = 2,30; $p < 0.001$). While ProcG was more active for PracFs and NovFs than their system-trained colleagues, this pattern was reversed for CtrlPanFs. There was a significant main effect of layoff ($F = 14.14$; d.f. = 1,15; $p < 0.005$), both groups carrying out nearly twice as many interventions at T_2 than T_1 . Overall, these results show that participants carried out more control actions at the second testing session, which is consistent with their improved control performance at T_2 .

5.3.2. Information sampling behaviour: Similar to the complex pattern found for MCA, multiple effects of training were also observed for information sampling behaviour. As table 2 shows, there was a propensity of SysG to make more use of the flow meters and the maintenance facility, both displays of high relevance for diagnostic activities. However, statistical tests failed to confirm significance of either effect. A significant effect of the layoff period was recorded for both displays but in different directions. While there was an increase in flow meter sampling at T_2 ($F = 4.74$; d.f. = 1,15; $p < 0.05$), there was a sharp reduction in interrogations of the maintenance facility for both groups ($F = 19.8$; d.f. = 1,15; $p < 0.001$). Whereas these two displays have primarily diagnostic functions, the history display is a dual function facility—not only does it have diagnostic properties, but also it represents a monitoring aid to identify system disturbances. The analysis revealed that after the layoff period, SysG participants sampled the history display less frequently (table 2). No such change was observed for ProcG. This means that the difference between the groups observed at T_1 increased at T_2 , giving a significant interaction ($F = 4.68$;

d.f. = 1,15; $p < 0.05$). There were also main effects of training ($F = 4.59$; d.f. = 1,15; $p < 0.05$) and layoff period ($F = 6.37$; d.f. = 1,15; $p < 0.05$) for this measure.

5.3.3. *System management strategies*: Since there is a close functional relationship between system monitoring and interventions, it is useful to relate the changes observed for information sampling behaviour and manual control activities. Control activities increased after the retention interval (in particular for ProcG) while information-sampling activities showed no overall increase. There was a tendency for SysG to increase the level of intervention between T_1 and T_2 , at the expense of monitoring, with increasing fault difficulty. No such pattern was observed for ProcG. Overall, the approach to system management, and the way this changes over time, shows considerable differences between the training groups.

5.4. *Subjective state measures*

The subjective state of the participants was measured at the end of each trial. Responses were made on 100-mm visual analogue scales for three variables: Effort, Anxiety and Fatigue. The data are presented in table 3. Little difference was found between training groups for mental effort expenditure and anxiety (for both $F < 1$). The data appeared to show that participants made more effort and were less anxious after the retention interval but again, the difference was not significant (for both $F < 1$). However, a main effect for fatigue was found with increased fatigue levels reported at T_2 ($F = 1.67$; d.f. = 1,15; $p < 0.05$). The observed main effect was due to a significant interaction between training and layoff ($F = 3.08$; d.f. = 1,15; $p < 0.05$) because ProcG showed an increase in reported fatigue levels from T_1 to T_2 while levels for SysG remained relatively stable.

5.5. *Mental model of operator*

The results for both explicit and implicit knowledge are presented in table 3. They indicate that system trained participants had a higher level of system knowledge at both testing times and for both types of knowledge ($F = 10.3$; d.f. = 1,15; $p < 0.01$). It was also found that the mental model improved over the retention interval for both training groups and both types of knowledge ($F = 4.8$; d.f. = 1,15; $p < 0.05$). No interaction was found ($F < 1$).

6. Discussion

In contrast to expectations, training had overall only a limited impact on performance, merely a few measures showing an effect. More marked effects were

Table 3. Self-reported subjective state measures (mean ratings on a 0–100 visual analogue scale) and assessment of system knowledge (mean score) at T_1 and T_2 .

	System group		Procedure group		T_1	T_2
	T_1	T_2	T_1	T_2	overall	overall
Effort ratings (0–100)	48.7	51.9	53.0	53.9	50.9	52.9
Anxiety ratings (0–100)	44.7	38.2	36.5	38.6	40.6	38.4
Fatigue ratings (0–100)	53.9	51.7	39.3	53.7	46.6	52.7
System knowledge (max score= 6)	2.6	3.3	0.7	1.8	1.6	2.5
Explicit (max score= 3)	1.2	1.6	0.5	1.1	0.8	1.3
Implicit (max score= 3)	1.4	1.7	0.2	0.7	0.8	1.2

observed for system management, notably for manual control activities, but to some extent also for information sampling strategies. The results also showed 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 but does not represent an insurmountable problem.

One of the obvious limitations of the study is the small sample size ($n = 17$). This was unfortunate but unavoidable, given the complex demands on participants, and the problems of maintaining the full sample over the 8 months between testing sessions. Because of this, analyses are necessarily carried out with rather low statistical power (between 0.12 and 0.52 for different analyses), so that there is a risk of failing to reject the null hypothesis when it is false. However, it is relevant to note that effect sizes (Cohen 1992) are, in any case, generally 'small' (Cohen's $d < 0.5$). The only two comparisons that show even moderate effect sizes are the reduced maintenance sampling at T_2 ($d = 0.58$) and the increase in system knowledge ($d = 0.56$). Thus, even if a more powerful design were employed, one would still wish to conclude that the effects of the long layoff in skill use are much less than would normally be expected. Furthermore, the overall effect of layoff for the most critical performance measure, management of system failures, is actually in the 'wrong' direction. If anything, control performance improves from T_1 to T_2 .

It was predicted that ProcG would be better at managing practised faults but that the more difficult faults (both novel and control panel failures) would be most effectively dealt with 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 understand better unfamiliar fault states and to deal with them more effectively. The results for parameter control failures were in the expected direction (but did not reach statistical significance), with SysG performing better on both of the more difficult fault states. In addition, SysG had a more complete mental model on both explicit and implicit knowledge. It is therefore surprising that SysG participants did not do better on the primary control task. There are well-known problems associated with predicting performance from different types of knowledge (e.g. Rowe *et al.* 1996), and Broadbent *et al.* (1986) and others (e.g. Buchner *et al.* 1995) have reported the dissociation between verbal system knowledge and displayed performance in control tasks. However, Rowe *et al.* have pointed out that the identification of meaningful action patterns can increase the predictive value of knowledge structures. This means focussing on those aspects of performance that are most closely related to the knowledge structures under investigation.

While these issues may also have played a role in the present study, further analyses pointed at other possible causes, too. It emerged that certain system management patterns were more frequent in SysG, which appear to have offset their initial advantages. System-trained participants set more subsystems to manual than they could manage, which resulted in unnecessary deviations from the target state because the control load was excessive. The difficulties operators have in managing an increasing number of manually controlled subsystems have been clearly demonstrated in some field studies with the CAMS task (Sauer *et al.* 1999a, c). Furthermore, it appeared that SysG participants intervened in situations where such actions were not entirely justified. This suggests that the more exploratory behaviour of SysG was, under some circumstances, counterproductive. Both patterns may have

been caused by an over-confidence of SysG participants in managing the system, since previous research has indicated that more confident operators make more use of manual control skills (Lee and Moray 1994). In hindsight, it would have been beneficial if one had included a measure of operator confidence since it would have enabled the hypothesis to be tested. It is recognized in training programmes that there is a fine line between instilling the necessary level of confidence and giving rise to over-confidence in trainees (and, similarly, the relation between trust, overtrust and mistrust; Muir 1988). The results demonstrate that even with a training method that produces a richer mental model, there is no guarantee that such undesired side-effects will not occur.

Despite the absence of effects of training on performance, considerable differences in system control and system monitoring were observed. This suggests that training substantially shaped the system management strategies used by operators. Information sampling and system interventions need to be examined together because of their close functional relationship. An increased fault difficulty (particularly the CtrlPanF condition) was found that led to a general shift from monitoring behaviour (e.g. reduced sampling of the history display) to more control activity. This is not unexpected since monitoring requirements are highest in automatic systems (Sheridan 1987) and system disturbances of increasing difficulty demand more interventions. Similar changes in the intervention/monitoring ratio have been demonstrated in previous research, as a function of sleep deprivation (Hockey *et al.* 1998). Interestingly, the present study showed that this shift was more pronounced for SysG, which may be indicative of their higher adaptability to changes in operational requirements. Marked changes in system management strategies have been observed in a number of studies, which showed that primary performance was largely maintained but this was only achieved by adopting a more suitable strategy (Hoc and Moulin 1994, Spérandio 1995). A typical example of these adjustments may be found in the context of process control, where operators move from open loop to closed loop control under increased workload (Bainbridge 1974).

An important aspect of the effectiveness of training is the long-term skill retention, as opposed to immediate performance gains. While complex continuous tasks of the kind used in the study are thought to be relatively less vulnerable to forgetting, one would generally expect some performance degradations after an 8-month layoff period. For the primary control task, while control errors did increase for practised faults after the layoff period, operators were actually better in dealing with the complex CtrlPanF problems. Models of skill decay (Rose 1989) would predict that PracFs would be less likely to show decrements because of the greater practice. 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 attain previous performance levels. Instead, surprisingly, PracF was the only fault type that showed a decrement. While the observation of such a decrement does not contradict current theories of skill acquisition, there is clearly a need to address the question of why performance deterioration was absent for NovF, and why performance for CtrlPanF improved.

Several hypotheses may be advanced for this unexpected finding. First, although considerable training was given before the first session, more difficult faults may have benefited from the additional learning opportunity afforded by the second session. Second, knowledge gains may have occurred following the mental model testing that followed the task at T_1 , which required participants to solve

problems concerning control panel failures. Although great efforts were made not to reveal solutions to participants, the session may have encouraged them to think more about these fault states and their possible control strategies. In addition, this session may have given them more confidence (Lee and Moray 1994) 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, without participants knowing the solution. The second explanation appears to be stronger than the first for the following reason. If there had been a learning process, one would have expected some performance improvements in the second and third hour of the first testing session. These however were not observed. Evidence for a more active coping with CtrlPanF at T_2 comes from the manual control action data. These indicated an overall increase in manual control activity for both groups, though it was much more pronounced for ProcG. This was in contrast to the perhaps very cautious approach to system control employed at T_1 , which displayed a stronger reliance on the automatic system. This strategy might have been the direct consequence of the procedure-based training method that discouraged an explorative approach towards the system.

A third possible explanation for the improved control performance for complex faults at T_2 is that participants became more selective in managing the overall task environment. They may have focused more strongly on the primary task at the expense of secondary tasks, as predicted by the compensatory control model. The calculation of a selectivity index (SI) allows for the test of this hypothesis (Hockey *et al.* 1998). First, performance differences as a function of retention interval ($T_2 - T_1$) are standardized for each performance variable, using a z -transformation. Performance improvements over the layoff period are indicated by a negative value (since the data are based on error scores). The selectivity hypothesis was tested by calculating SI for RT and parameter control failures ($Z_{RT} - Z_{PCF}$). If SI has a positive value, it shows that the performance on the primary control task is relatively better than on the secondary task after the layoff. For example, a positive SI will result if there are no changes over the layoff period for RT (i.e. $Z_{RT} = 0$) while control performance improves (i.e. $Z_{PCF} =$ negative value), or if RT gets much worse than control performance. The analysis showed that participants appeared to become overall more selective at T_2 , though the difference was not statistically significant ($F < 1$). However, a significant interaction between layoff and fault type was found ($F = 4.96$; d.f. = 3,45; $p < 0.005$), indicating that a significant increase in selectivity occurred only for the most demanding fault condition (LSD-test $p < 0.001$). This would be in line with the performance improvements for CtrlPanFs found at T_2 . No significant effect of training emerged for SI but, not surprisingly, analysis of variance confirmed that with increasing fault difficulty, participants increased their focus on the control task ($F = 9.42$; d.f. = 3,45; $p < 0.001$). It seems likely that the effects of layoff on the primary task are the result of both this change in compensatory control and the above argument concerning the after-effects of the mental model session. An awareness that the complex faults may be manageable may be one of the reasons why more resources are devoted to them in the retest.

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 (Farr 1987, Annett 1989, Rose 1989). The improved control performance is

underpinned by an improvement of the participants' mental model after the layoff interval. Although the observed improvement of the mental model is consistent with the results from the performance measures, it is at odds with findings that consider mental models as temporally unstable and as being subject to a trend of simplification (Norman 1983). In contrast to the tracking skills, the recollection of diagnostic rules was affected by non-practice, resulting in longer identification times while accuracy was maintained. This effect occurred only for ProcG, suggesting that their presumed strategy of remembering procedures rather than using their system knowledge may have been less resistant against time-induced forgetting. This is supported by evidence in the literature, which suggested that a set of procedures is difficult to retain over time (Hagman and Rose 1983, Rose 1989). However, unlike in these studies, the present procedure-trained 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. In hindsight, it would have been useful to collect data on the use of the fault finding guide, for example, by providing a software-embedded version of it.

As predicted by the theoretical framework of the study, decrements of secondary task performance as a function of fault difficulty were found. In the context of resource theory (Navon and Gopher 1979, Wickens 1992), one may argue that, with increasing difficulty of the fault state, the cognitive requirements for the primary task increased, which resulted in fewer resources being available for the secondary tasks. This pattern emerged very clearly for prospective memory and reaction time.

7. Conclusion

The pattern of results from the present experimental work was largely in line with the predictions of the model of compensatory control mechanisms. Primary task performance was largely maintained. The observed impairment of secondary task performance (through increased demands as a function of fault states) may be interpreted as being the result of a reduction in residual capacity. This means a redirection of resources occurred, away from the secondary tasks, to increase the available resource pool for the primary task. In addition to secondary task performance decrements, increases in demands led to adaptive changes in information sampling behaviour and intervention strategies.

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 SysG had a better mental model and even demonstrated that in control performance, it did not pay off in terms of achieving better performance in the measures used. The analysis of the system management approach also corroborated the finding that the training methods resulted in different knowledge structures. However, due to non-optimal system management the SysG did not benefit from their initial advantage. This demonstrates that even with a better training method, there is no guarantee that undesired side-effects of this method will not offset its advantages. This reiterates the point that great care needs to be taken in the design of training, in particular, concerning those aspects of the training programme that go beyond the knowledge base, such as confidence, trust and risk-taking.

Acknowledgements

Part of the work was funded by the Long Term Programme Office of the European Space Agency (Contract no. RFQ-3-7873/93/F/TB). The authors are very grateful for their practical and financial support. In particular, thanks are due to Mr A. Bichi (ESA-LTPO) for constant support and encouragement over many years. Finally, they thank Anthony Dobson and Suzanne Al-Sarraf for assistance with parts of the experimental work.

References

- ANNETT, J. 1989, Trained skilled performance. In A. M. Colley and J. R. Beech (eds), *Acquisition and Performance of Cognitive Skills* (Chichester: Wiley).
- ARTHUR, W., BENNETT, W., STANUSH, P. L. and MCNELLY, T. 1998, Factors that influence skill decay and retention: a quantitative review and analysis, *Human Performance*, **11**, 57–101.
- 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 & Francis).
- BRANDIMONTE, M., EINSTEIN, G. O. and MCDANIEL, M. A. 1996, *Prospective Memory: Theory and Applications* (Hillsdale: Erlbaum).
- 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.
- BUCHNER, A., FUNKE, J. and BERRY, D. 1995, Negative correlations between control performance and verbalizable knowledge: indicators for implicit learning in process control tasks?, *Quarterly Journal of Experimental Psychology*, **48**, 166–187.
- COHEN, J. 1992, A power primer, *Psychological Bulletin*, **112**, 155–159.
- 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 & Francis).
- DANIELLOU, F. 1986, *L'opérateur, la vanne, l'écran: L'ergonomie des salles de contrôle* (Montrouge: ANACT).
- DE KEYSER, V. 1988, How can computer-based visual displays aid operators? In E. Hollnagel, G. Mancini and D. D. Woods (eds), *Cognitive Engineering in Complex Dynamic Worlds* (London: Academic Press).
- 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).
- FARR, M. J. 1987, *The Long-term Retention of Knowledge and Skills: A Cognitive and Instructional Perspective* (New York: Springer).
- HAGMAN, J. D. and ROSE, A. M. 1983, Retention of military tasks: a review, *Human Factors*, **25**, 199–213.
- 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. 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: Oxford University Press).
- HOCKEY, G. R. J. 1997, Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework, *Biological Psychology*, **45**, 73–93.
- HOCKEY, G. R. J., WASTELL, D. and SAUER, J. 1998, Effects of sleep deprivation and user interface on complex performance: a multilevel analysis of compensatory control, *Human Factors*, **40**, 233–253.
- LEE, J. D. and MORAY, N. 1994, Trust, self-confidence and operators' adaptation to automation, *International Journal of Human-Computer Studies*, **40**, 153–184.

- MORRIS, N. M. and ROUSE, W. B. 1985, Review and evaluation of empirical research in troubleshooting, *Human Factors*, **27**, 503–530.
- 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).
- NAVON, D. and GOPHER, D. 1979, On the economy of the human processing system, *Psychological Review*, **86**, 254–255.
- NORMAN, D. A. 1983, Some observations on mental models. In D. Genter and A. L. Stevens (eds), *Mental Models* (Hillsdale: Lawrence Erlbaum).
- 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.
- 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).
- ROUSE, W. B. and MORRIS, N. M. 1986, On looking into the black box: prospects and limits in the search for mental models, *Psychological Bulletin*, **100**, 349–363.
- ROWE, A. L., COOKE, N. J., HALL, E. P. and HALGREN, T. L. 1996, Toward an on-line knowledge assessment methodology: Building on the relationship between knowing and doing, *Journal of Experimental Psychology: Applied*, **2**, 31–47.
- SARTER, N. B., WOODS, D. D. and BILLINGS, C. E. 1997, Automation Surprises. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 1926–1943.
- SAUER, J., HOCKEY, G. R. J. and WASTELL, D. 1999a, Maintenance of complex performance during a 135-day spaceflight simulation, *Space, Aviation, and Environmental Medicine*, **70**, 236–244.
- SAUER, J., HOCKEY, G. R. J. and WASTELL, D. 1999b, Performance evaluation in analogue space environments: adaptation during an 8-month Antarctic over-wintering expedition, *Space, Aviation, and Environmental Medicine*, **70**, 230–235.
- SAUER, J., WASTELL, D. and HOCKEY, G. R. J. 1997, Skill maintenance in extended spaceflight: a human factors analysis of space and analogue work environments, *Acta Astronautica*, **39**, 579–587.
- SAUER, J., WASTELL, D. and HOCKEY, G. R. J. 1999c, Multiple-task performance on a computer-simulated life support system during a space mission simulation, *Acta Astronautica*, **44**, 43–52.
- SAUER, J., WASTELL, D. and HOCKEY, G. R. J., 2000, Using micro-worlds to simulate highly-automated work environments: the case of the Cabin Air Management System, *Computers in Human Behavior*, **16**, 45–58.
- 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. 1987, Supervisory control. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 1295–1327.
- 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), 649–656.
- STAMMERS, R. 1996, Training and the acquisition of knowledge and skill. In P. Warr (ed.), *Psychology at Work* (London: Penguin), 73–94.
- SWEZEY, R. W. and LLANERAS, R. E. 1997, Models in training and instruction. In G. Salvendy (ed.), *Handbook of Human Factors* (New York: Wiley), 514–577.
- WICKENS, C. D. 1992, *Engineering Psychology and Human Performance* (Columbus: Merrill).
- 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).
- WOODS, D. D. 1988, Commentary: Cognitive engineering in complex and dynamic worlds. In E. Hollnagel, G. Mancini and D. D. Woods (eds), *Cognitive Engineering in Complex Dynamic Worlds* (London: Academic Press).
- WOODS, D. D., O'BRIAN, 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).