

The effects of heuristic rule training on operator performance in a simulated process control environment

Jürgen Sauer^{a*}, Dina Burkolter^b, Annette Kluge^b, Sandrina Ritzmann^b and Kerstin Schüler^b

^a*Department of Psychology, University of Fribourg, Rue de Faucigny 2, CH-1700, Fribourg, Switzerland;* ^b*Research Institute for Organizational Psychology, University of St. Gallen, Varnbühlstrasse 19, CH-9000, St. Gallen, Switzerland*

In complex work environments, the occurrence of novel system states represents a particular challenge for the design of training. This article is concerned with the use of heuristic rules to prepare operators for the management of unfamiliar fault states. An experiment was carried out to examine the effects of heuristic rule training on operator performance and system management behaviour. Thirty-nine trainee operators from the chemical industry took part in the study. They were trained for 4 h on a PC-based simulation of a process control task. Operators in the experimental group received training on heuristic rules while operators in the control group did not. One week later the operators participated in a 70-min testing session. While the results showed that heuristic rules training led to better diagnostic performance, it was also associated with increased operator fatigue and impaired secondary task performance. The implications of the findings for using heuristic rule training are discussed.

Keywords: training; heuristic rule; process control; performance

1. Introduction

1.1. Training in complex systems

The design of training has long been considered to be of high importance for a smooth operation of complex human-machine systems. Over recent years, the importance of learning has even increased due to more changes in work environments (e.g. Warr 2002). While there have been improvements in the effectiveness of training, various aspects of modern technical work systems continue to present challenges to the design of training. First, technical systems can be highly complex (due to opaqueness, time lags, etc.), which makes it difficult for operators to learn how to operate these. This even applies to the completion of routine operations (i.e. familiar and well-practised task scenarios), which can be anticipated by training instructors. Second, routine operations may become more challenging under increased task load. Although the task scenarios are still familiar to the operator, the temporal accumulation of events (e.g. air traffic controller has to manage simultaneously 12 instead of six aircraft) makes it difficult to maintain performance levels. Third, task completion has to be carried out under the presence of

external stressors, which reduces the amount of cognitive-energetical resources available to the operator. For example, noise serves as a distracter, which diverts resources from the main tasks. Fourth, the operator may be faced with unfamiliar task scenarios, which the operator has never encountered before and hence requires some adaptive response to the novel situation. Working conditions will be even more aggravated when more than one of these factors is present (e.g. unfamiliar task scenario under noise). Since the operational conditions outlined may differ in terms of the kind of training needed, it provides an indication of the difficulties of selecting and designing effective training methods.

The body of literature on training design for complex systems is quite extensive (for an overview, see, for example, Salas *et al.* 2006). A plethora of training methods has been developed, including drill and practice (e.g. Gopher *et al.* 1989), procedure-based training (e.g. Patrick and Haines 1988), knowledge-based training (e.g. Hockey *et al.* 2007), error training (e.g. Lorenzet *et al.* 2003), cognitive skills training (O'Brien and O'Hare 2007) and cognitive apprenticeship (e.g. Collins *et al.* 1989). While all of these training methods have their merits, their relative effectiveness depends on the operational circumstances under which they are being used (Kluge *et al.* in press). The present article is primarily concerned with the preparation of operators for unfamiliar fault states for which no established procedures are available. For these operational requirements, a knowledge-based training approach may be beneficial since it provides operators with a deep understanding of the technical system, which allows a flexible response to cope with novel problem situations (Hockey *et al.* 2007). This includes, for example, an understanding of the relationships between different elements and parameters of the system (e.g. valves, controls, pipe pressure), similar to the kind of high-level knowledge that the designer of the system has, though clearly not at the same level of detail and comprehensiveness. Due to the high-level knowledge conveyed by the training method, considerable cognitive demands are placed on the operator, requiring an appropriate level of skill and expertise for successful information processing. Therefore, this may be a sensible approach when novice operators are to be trained. For novices (e.g. trainee operators, apprentices), heuristic rule training (HRT) may be a promising alternative since it places lower demands on operator skills and abilities.

1.2. Heuristic rule training

When the term 'heuristic' is referred to in the psychological research literature, it typically concerns human decision-making and a number of biases that influence that cognitive process, such as availability heuristics (Tversky and Kahneman 1974). In the field of cognitive psychology, heuristics are also referred to as rules-of-thumb (Eysenck and Keane 2005), while, in creativity research, heuristics have been employed for the purpose of enhancing creative thinking (e.g. Amabile 1983).

In the field of work psychology and ergonomics, heuristic rules are considered to be cognitive aids that may be used to support planning, implementation and control of complex work activities (Frieling and Sonntag 1999, Sonntag and Schaper 2006). They represent short and precise instructions given to the operator to generate and evaluate actions that may help to deal with an unfamiliar operational scenario, to which established procedures cannot be successfully applied. Similarly, Salas *et al.* (2006) referred to heuristics as job aids that guide the operator through a thought process to find the best solution to a problem. This makes it very distinct from procedural aids, which provide a

series of steps in a predefined order. However, other work in the research literature defined heuristic rules in a slightly different way, de-emphasising their generic nature. For example, the 'heuristic rules' for fault diagnosis employed by Linou and Kontogiannis (2004) have rather the characteristics of a set of procedures that can be used as a decision aid (e.g. 'Check for flow problems in overhead system'). In the present article, the former definition of heuristics is adopted, which emphasises their utility for unfamiliar operational scenarios.

Finer distinctions may be made between different types of heuristic rules (Hacker and Skell 1993). They can differ in their level of specificity, ranging from highly generic rules (e.g. 'Compare current state to target state!') to very specific ones (e.g. 'Compare technical drawing to work piece!'). Furthermore, heuristic rules may be of a motivational nature (e.g. 'Take action before it is too late!') or they may aim to structure the actions to be carried out (e.g. 'Check that the steps to be taken are in the correct order!').

In German work psychology, there is a small body of literature that examined the effectiveness of heuristic rules in training (e.g. Volpert *et al.* 1984, Sonntag and Schaper 1988, Schoppek 1997). Sonntag and Schaper (1988) showed that apprentices trained on heuristic rules in the work environment of complex automated production plants worked more systematically and committed fewer diagnostic errors than a control group. These improvements in performance were generally attributed to the use of improved meta-cognitive strategies (i.e. operators' knowledge about their own cognitive processes and strategies). Benefits for HRT were not only found for diagnostic tasks but also for system control tasks, as a study by Volpert *et al.* (1984) has indicated. Using a simulated process control environment that modelled a bread factory, it showed that participants trained on heuristic rules performed better on a system control task than a control group, which was attributed to the more frequent sampling of task-critical information. Other work employing a simulated manufacturing task showed similar benefits of HRT for the adoption of a more systematic approach to task management, which was, however, not paralleled by improved control performance (Schoppek 1997). Another experiment conducted in the application area of process control only provided a single heuristic rule ('Always select the response level half-way between the current production level and the target level, and you will get as close to the target level as possible!') to the operators (Stanley *et al.* 1989). Being considered to be the golden rule for successful system management, the findings confirmed its effectiveness since it led to better performance than the control group.

The literature review revealed that HRT was effective in improving performance in fault diagnosis as well as in system control but the positive effects were not consistently observed. This may be due to differences in the use of heuristic rules with regard to their specificity and the way in which they were derived (e.g. 'Was a task analysis conducted?'). Interestingly, none of the studies examined fault diagnosis and system control together in a multiple-task paradigm, nor was any other combination of tasks used. In the studies reviewed that were conducted in the field, HRT was utilised as a very basic method of preparing operators for unfamiliar operational scenarios. Therefore, HRT offers particular advantages in terms of low requirements for skill levels and cognitive abilities. For this reason, it may also be more suitable for novice trainees than experts. HRT conveys context-independent strategies in explicit ways, which are often already implicitly used by experienced operators. Finally, the low demands of HRT on training time compared to alternative training methods need to be taken into consideration in a cost-benefit analysis.

1.3. The present study

This study aimed to examine the utility of heuristic rules in training for familiar and novel scenarios. Familiar scenarios were those that were extensively practised in training while novel scenarios had not been encountered before.

The work was carried out in the field of process control by using a generic computer simulation of a complex process control environment. While previous research on heuristic rules has either focused on fault diagnosis or on system control, the present study aimed to include both aspects of performance. This made it possible to examine the effectiveness of heuristic rules for either task activity, which are characterised by considerable differences. System control focuses on a forward flow of events (i.e. ‘What causes what?’), whereas fault diagnosis is more concerned with the reverse pattern (i.e. ‘What is caused by what?’). Since operators find it generally less difficult to do forward reasoning than backward reasoning (Wickens and Hollands 2000), there may be a greater need to support operators in fault diagnosis than in system control.

Making use of a multiple task environment rather than employing a single task setting (as in previous studies) reflects more adequately the complexity of real work activities. Apart from the more realistic modelling of real work and the ensuing increase in ecological validity, the inclusion of secondary tasks has the added benefit of allowing workload assessment measures to be taken.

The work examined the effectiveness of heuristics provided as an add-on to a standard training approach. This essentially corresponds to a procedure-based training regime, which represents the basis of many training approaches (cf. Morris and Rouse 1985). It involves the provision of a set of procedures, which can be employed by operators to deal with a wide range of operational situations.

In the present study, data on operator characteristics (e.g. mental abilities, work-relevant personality variables, working style) were also collected to assess their predictive qualities for determining operator performance. The findings of these analyses are reported in a separate article (Burkolter *et al.* under review).

Based on the literature review, two main research assumptions were formulated: (a) HRT would be more effective than a standard training approach for both primary task parameters (i.e. fault diagnosis and system control). The effectiveness of HRT compared to the standard training would be more pronounced for novel faults than for familiar faults states; (b) HRT would encourage operators to make more use of meta-cognitive strategies than the standard training approach.

2. Method

2.1. Participants

In all, 39 participants (51.3% female) took part in the study. All of them were trainee operators of different Swiss chemical companies, which make use of complex process control environments in their production systems. Their ages ranged from 16 to 22 years (mean 18.1).

2.2. Design

A 2 × 4 mixed experimental design was employed. Training as a between-participants variable was varied at two levels (heuristic vs. basic) and fault type as a within-participants

variables was varied at four levels (fault-free, practised faults, novel faults, complex faults). Details of these are given below.

2.3. The task environment

In the present study a computerised simulation of a generic process control environment was employed, modelling life support systems onboard a spacecraft. As the task environment, called Cabin Air Management System (CAMS), has been described in detail elsewhere (Sauer *et al.* 2000b), it is merely briefly summarised here.

CAMS comprises a number of automatic controllers that maintain five system parameters (O_2 , CO_2 , cabin pressure, temperature, humidity) within a specified range. Figure 1 shows the main interface of CAMS. The topographical layout of the system is displayed, with its main features, such as warning system, history display, flow meter readings and parameter control panel.

The operator is required to complete four tasks, of which two were defined as primary tasks and two as secondary tasks. (a) The first primary task was to identify and repair any system failure. Two aspects of diagnostic performance were measured: fault identification time and diagnostic errors; (b) the second primary task was to maintain a stable system state at all times. This was achieved by monitoring the safe functioning of the automatic controllers and by adopting manual system control if required; (c) system alarm

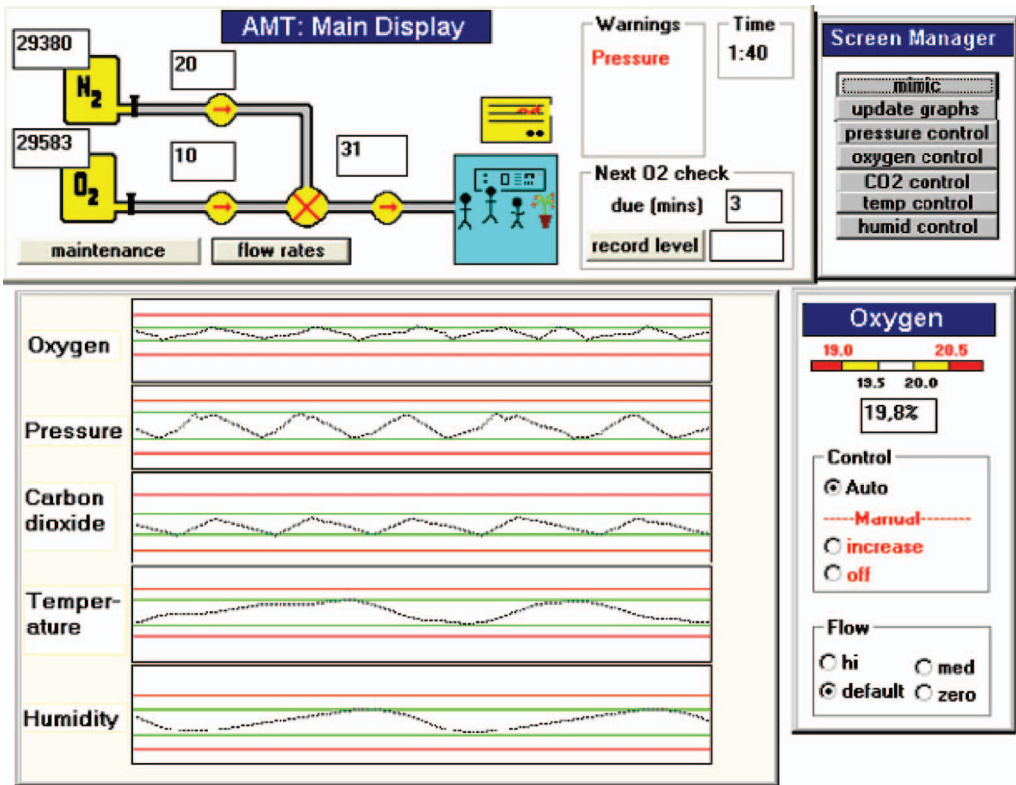


Figure 1. Main display of the Cabin Air Management System.

acknowledgement referred to a secondary task, which involved responding to system alarms as soon as they were displayed on the screen. This task provided a measure of reaction time; (d) the other secondary task involved the completion of regular checks of O₂ tank levels, corresponding to a prospective memory task (i.e. to remember to complete an action at a specified time in the future). In the present case, this was done by recording O₂ tank levels at 3-min intervals.

All system interrogations and interventions carried out by the operator are recorded by CAMS. This allowed the measurement of different system management behaviours, such as flow meter viewings, history display sampling, maintenance facility viewings and manual control actions.

CAMS permits the collection of subjective state measures at 30-min 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 ratings of mental effort, anxiety and fatigue by moving the slider along the scale with the mouse. The number of items was deliberately kept to a minimum to reduce task disruptions, which becomes a pertinent issue if the items are presented repeatedly during a performance test. The use of one-item scales is considered to be acceptable if the items are unambiguous and capture the main concept (Wanous *et al.* 1997).

2.4. Training

Participants were randomly allocated to one of the training conditions, with 19 participants receiving HRT while 20 participants were given basic training (BT). In both groups, total training time was approximately 4 h, which was separated by a 10-min and a 20-min break. HRT and BT followed the same procedure, except for three units (totalling 25 min) in which the heuristic rules were presented to the HRT-group while the BT-group carried out a different activity (see 2.4.1). In the first training block the participants were introduced to CAMS and its features, to the primary and secondary tasks as well as to the fault-finding guide by means of a presentation and short exercises. In the next training blocks the following five system faults were presented and then practised by the participants: nitrogen valve auto failure, leak in O₂ valve, blocked O₂ valve, cooler set point failure and CO₂ scrubber ineffective. The training sessions typically took place in groups of nine or 10 participants, with each group being supervised by two instructors. Participants worked on individual computers during training. A fault-finding guide was given to participants, which they could consult during training and the experimental testing session.

2.4.1. Basic training

This form of training corresponded to the standard training approach used in safety-critical industries to ensure that safe and efficient ways of system management are adopted by the operator. It mainly involved teaching procedures to the trainees. During training, the importance was stressed to follow the set of procedures for fault diagnosis and system stabilisation. All procedures were available in a fault-finding guide. While the HRT-group was introduced to the heuristic rules, the participants in the BT condition were given information about the required recruitment criteria for astronauts in the European space programme and the psychological demands placed on them. This experimental condition represented the control group.

2.4.2. Heuristic rule training

The HRT-group received the same procedure-based training regime as the control group. In addition, they were introduced to a set of six heuristics (see Figure 2). The first three rules were designed to support the operator during fault diagnosis, whereas the last three referred to system control activities. The teaching of the heuristic rules consisted of three units (totalling 25 min), which were integrated into the basic training regime. In the first unit (15 min), the general advantages of using heuristic rules were pointed out, followed by presentation of the six heuristic rules. Their utility in the context of CAMS management was explained and, with a view to activating the knowledge structures of the participants, they were subsequently asked to voice their deliberations about the applicability of the heuristic rules. In the second and third unit (5 min each) the use of the rules was repeated and more extensively practised. If the rules were not appropriately used, participants were made aware of this by the instructors. Figure 2 also shows typical examples of errors that are to be avoided by following the heuristic rule. The rules derived for HRT were based on a hierarchical task analysis of CAMS and an error analysis on the data of previous CAMS studies (Burkolter *et al.* 2007). The error analysis was carried out by the cognitive reliability and error analysis method, called CREAM (Hollnagel 1998).

2.5. Fault scenarios

Fault type was varied as a within-subjects variable at four levels. In the fault-free condition, the automatic system required no operator intervention. Practised faults referred to fault states that had been extensively practised during the training sessions. Novel faults were of the same quality as practised faults but had not been encountered during training. Complex faults represented a different class of faults because they had features that were not practised in training. In addition to being unknown to participants,

Heuristic rule	Example of error
– Remember to carry out diagnostic checks in the right order!	– Operator cut corners in fault diagnosis and repair, resulting in premature (and wrong) identification of cause of fault state.
– Do not forget to check whether your actions actually achieved their goals!	– Operator repaired system fault but did not check afterwards whether repair was actually successful (i.e. was diagnosis correct?).
– Beware confusing one fault state with another!	– Operator did not check all criteria before making a diagnosis, resulting in the selection of a similar but incorrect system fault.
– Take action if the system becomes unstable!	– Although system state had become unstable, the operator waited for a long time before taking over manual control (because operator expected that the automatic controller will recover the system).
– Take action even if the situation is unfamiliar to you!	– Operator did not take any action because several subsystems of CAMS failed simultaneously (e.g. control panel and maintenance facility).
– Consider how the different system elements are related to each other!	– Operator had a tunnel vision of problem state and did not think of alternative options to stabilise the system.

Figure 2. Heuristic rules (derived from a hierarchical task analysis and a cognitive reliability and error analysis; Burkolter *et al.* 2007) and examples of the kind of error that each rule aims to prevent. CAMS = Cabin Air Management System.

this type of fault could not be repaired (because the maintenance facility was unavailable) and the most obvious strategy of controlling it could not be used (manual control panel most needed was unavailable), requiring an indirect way of dealing with the fault (e.g. rather than switching the cooler on to deal with increasing cabin temperature, the participant needed to increase the supply of low-temperature N₂).

2.6. Assessment of meta-cognitive activities and mental model

A six-item questionnaire was developed to measure the meta-cognitive activities that the heuristic training was expected to stimulate. Three of the items referred to fault diagnosis ('I made sure not to mix up different system faults', 'I adhered to the correct sequence of actions in the fault-finding guide' and 'When making adjustments, I regularly examined their effects on the system'). The three other items concerned system control ('I took action, as soon as the system threatened to become unstable', 'I thought about how the system elements were related to each other' and 'I took action when even unexpected system faults occurred'). All items had a 6-point Likert scale (strongly disagree to strongly agree).

The explicit mental model (i.e. declarative knowledge) was measured by a questionnaire, which had been used in previous studies with the CAMS environment (e.g. Sauer *et al.* 2000a, 2006). This was considered to be an efficient instrument to capture the mental model since it was considerably less time-consuming to use than employing alternative methods such as verbal protocols. Previous work showed that this approach provided similar results, such as a verbal protocol analysis (Sauer *et al.* 2000a). The instrument comprised 12 three-alternative multiple choice items (e.g. 'What happens to the cabin temperature when nitrogen is pumped into the cabin?' 'Increase – decrease – no effect'), with each item requiring a short explanation why the given answer was correct. Furthermore, the questionnaire contained three open questions (e.g. 'Please explain which components or processes have an impact on O₂ level in the cabin and describe the direction of the relationship!'). Participants were able to achieve a score of 1 for each multiple-choice item (0.5 points for selecting the correct answer and another 0.5 points for giving the correct explanation) and a maximum of 3 points for each open question, resulting in a maximum possible score of 21 points. The scores achieved in the present sample ranged from 2.5 to 16.5 points.

2.7. Procedure

The testing session was held in the same group of participants, who were trained together. Participants were tested for 70 min, during which they encountered fault types with the following frequencies: three practised faults; three novel faults; and two complex faults. The order and duration of faults was balanced across fault types. Participants were unaware of the order, type and onset of faults. A typical set-up of a 35-min testing (2100 s) period is illustrated by the following example: fault free (0–129 s); practised fault (130–549 s); fault free (550–619 s); novel fault (620–1049 s); fault free (1050–1099 s); practised fault (1100–1469 s); fault free (1470–1559 s); complex fault (1560–2100 s).

To ensure that the heuristic rules were easily remembered, they were made visible to the participants in three different ways: (a) an index card (sized A5) was placed adjacent to the participant's computer screen; (b) three posters (sized A3 each) were hung on the wall; (c) a projector was used to display a power point slide on the wall (sized about 6 m²). After task completion, participants were presented with the questionnaires measuring meta-cognitive activities and the mental model.

3. Results

For some of the measures, the datasets of several participants needed to be removed from the data analysis. This was the case for flow meter sampling, where one participant (BT-group) was excluded because he sampled the flow meters more than twice as frequently as the participant with the second highest score, suggesting a system management strategy that consisted of having the display information available at all times. For the subjective state measures, one participant (BT-group) was excluded because he only made ratings at the extreme end of the scale, suggesting that he had not fully understood the purpose of the subjective state measures. If the assumption of sphericity was violated in the ANOVA, degrees of freedom (df) were corrected using Greenhouse-Geisser estimates of sphericity. Furthermore, wherever post-hoc tests were employed, the Bonferroni adjustment was used. An alpha-level of 0.05 was chosen but marginally significant effects (i.e. at an alpha-level of 0.1) were also reported. To indicate the practical significance of the effect, partial Eta-square is reported (or r for t-tests), where appropriate, as a measure of effect size.

3.1. *Meta-cognitive activities and mental model*

The analysis of the questionnaire results indicated slightly higher meta-cognitive activities for the HRT than the BT-group (5.0 vs. 4.6) on a scale ranging from 1 to 6. This difference between groups was marginally significant ($t = 1.88$; $df = 37$; $p = 0.066$; $r = 0.30$). An analysis of meta-cognitive activities, carried out separately for heuristic rules concerning fault diagnosis and system control, showed a similar pattern. Although the effects observed were not strong, the results indicate that the experimental treatment was successfully implemented.

Concerning the quality of the explicit mental model of the participants (i.e. knowledge that can be verbalised), there was little difference in operator knowledge between training groups (HRT: 9.6; BT: 10.4). A t-test confirmed that this difference was not significant ($t < 1$).

3.2. *Operator performance*

3.2.1. *Fault diagnosis (primary task)*

Diagnostic performance was measured by two parameters: accuracy and speed (see Table 1). The results showed that HRT improved diagnostic accuracy ($F = 5.23$; $df = 1, 36$; $p < 0.05$; $\eta^2_{\text{partial}} = 0.13$) but did not affect diagnostic speed ($F = 1.52$; $df = 1, 36$; NS). Not surprisingly, diagnostic performance was better for practised faults than for novel faults, observed for both parameters (accuracy: $F = 5.46$, $df = 1, 36$, $p < 0.05$, $\eta^2_{\text{partial}} = 0.13$; speed: $F = 29.1$, $df = 1, 36$, $p < 0.001$, $\eta^2_{\text{partial}} = 0.45$). Neither diagnostic accuracy nor speed showed an interaction between training and fault type (both $F < 1$).

3.2.2. *System control (primary task)*

The fault-free condition was not included in the analysis since only very few control errors were recorded under that condition (see Table 1). In contrast to diagnostic accuracy, the analysis revealed a marginally significant effect of training for system control performance, with the HRT-group performing worse than the BT-group ($F = 3.48$; $df = 1, 36$; $p = 0.07$; $\eta^2_{\text{partial}} = 0.09$). Furthermore, there was a significant interaction between

Table 1. Operator performance as a function of training and fault type.

	Heuristic rule training	Basic training	Overall
Incorrect fault diagnoses (%)	37.0 (22.5)	53.3 (21.3)	
Practised faults	29.6 (32.1)	48.3 (27.5)	39.5 (30.8)
Novel faults	44.4 (25.3)	58.3 (23.9)	51.8 (25.3)
Fault identification time (s)	295.0 (57.5)	317.5 (54.7)	
Practised faults	266.8 (76.4)	279.4 (67.2)	273.4 (71.0)
Novel faults	323.2 (64.6)	355.6 (62.3)	340.2 (64.6)
System control failures (%)	8.5 (3.5)	6.9 (1.6)	
Fault free	0.2 (0.4)	0.2 (0.4)	0.2 (0.4)
Practised faults	2.5 (1.8)	2.9 (3.3)	2.7 (2.7)
Novel faults	9.5 (5.1)	8.0 (3.0)	8.7 (4.2)
Complex faults	21.6 (8.2)	16.3 (3.3)	18.8 (6.6)
Reaction time (s)	2.4 (0.54)	2.1 (0.50)	
Fault free	1.9 (0.53)	1.7 (0.61)	1.8 (0.57)
Practised faults	2.3 (0.63)	2.1 (0.73)	2.2 (0.68)
Novel faults	2.7 (0.79)	2.3 (0.65)	2.5 (0.72)
Complex faults	2.6 (1.07)	2.3 (0.71)	2.4 (0.93)
Prospective memory failures (%)	23.7 (23.2)	16.6 (10.4)	
Fault free	2.8 (11.8)	5.0 (15.4)	3.9 (13.7)
Practised faults	21.4 (14.9)	13.6 (17.0)	17.3 (16.3)
Novel faults	23.8 (19.0)	20.7 (18.2)	22.2 (18.4)
Complex faults	33.3 (24.3)	19.2 (12.4)	25.9 (20.0)

Note: Standard deviation is shown in parentheses.

training and fault scenario ($F = 5.77$; $df = 1, 9$, 68.7 ; $p < 0.01$; $\eta^2_{\text{partial}} = 0.14$). This was because BT led to better performance for novel faults (post-hoc test: $p < 0.05$) and complex faults than HRT (post-hoc test: $p < 0.05$), whereas no differences between training groups were observed for practised faults and the fault-free condition. As expected, there was a main effect of fault scenario, with more control errors being observed with increasing scenario difficulty ($F = 234.4$; $df = 1, 9$, 68.7 ; $p < 0.001$; $\eta^2_{\text{partial}} = 0.87$). Post-hoc tests for the main effect of fault scenario indicated significant differences between all conditions ($p < 0.001$).

3.2.3. Reaction time (secondary task)

Monitoring of system alarms was slightly better in the BT-group than for the HRT-group, as the reaction time measure demonstrates (see Table 1). This difference was marginally significant ($F = 3.74$; $df = 1, 36$; $p = 0.061$; $\eta^2_{\text{partial}} = 0.09$). No interaction between training and fault scenario was found ($F < 1$). As expected, reaction time slowed down as fault scenarios became more difficult ($F = 10.4$; $df = 3, 108$; $p < 0.001$; $\eta^2_{\text{partial}} = 0.23$). Post-hoc tests confirmed that only the fault-free condition was different from the other three ($p < 0.01$).

3.2.4. Prospective memory (secondary task)

Similar to reaction time, slight performance decrements were observed for the HRT condition compared to the BT control group (see Table 1). This manifested itself in the form of an increased number of omitted tank level recordings, that is, operators forgot to carry out the prospective memory task. The difference between groups was marginally

significant ($F = 3.47$; $df = 1, 36$; $p = 0.071$; $\eta^2_{\text{partial}} = 0.09$). An analysis of the effect of fault scenario confirmed the decreasing performance with increasing scenario difficulty ($F = 13.5$; $df = 3, 108$; $p < 0.001$; $\eta^2_{\text{partial}} = 0.27$). As was the case for reaction time, post-hoc tests indicated that only the fault-free condition was significantly different from the other three conditions ($p < 0.01$). No interaction between training and fault scenario was observed ($F = 1.74$; $df = 3, 108$; NS).

3.3. Subjective state

The data for the system-embedded subjective state measures are presented in Table 2. The analysis revealed no effects of training for effort expenditure ($t < 1$) and anxiety ($t < 1$). In contrast, fatigue ratings were considerably higher for HRT than for BT, confirmed by a significant t-test ($t = 2.15$; $df = 35$; $p < 0.05$; $r = 0.34$).

3.4. System management behaviour

The data for information sampling behaviour and control action were not available for separate fault states. The analysis therefore examined only the effects of training, using a simple t-test. The data are presented in Table 3. The data for history display sampling show that the HRT-group sampled the history display less frequently than the BT-group. This difference was significant ($t = 2.36$; $df = 20.14$; $p < 0.05$; $r = 0.47$). For flow meter viewings, no such difference was found ($F = 1.37$; $df = 1, 35$; NS). If anything, the HRT-group sampled this information source somewhat more frequently than the BT-group. Finally, the analysis revealed no difference between training groups for the number of control actions completed ($t < 1$).

4. Discussion

The main finding of the study was that HRT was effective in improving diagnostic accuracy but showed no benefits for system control performance. There were some indications of HRT leading to unintended side effects in the form of higher fatigue ratings

Table 2. Subjective operator state as a function of training and fault type.

	Heuristic rule training	Basic training
Mental effort (0–100)	43.0 (23.0)	45.4 (24.8)
Anxiety (0–100)	38.7 (21.8)	35.8 (25.3)
Fatigue (0–100)	53.6 (26.9)	36.7 (20.8)

Note: Standard deviation is shown in parentheses.

Table 3. System management behaviour as a function of training and fault type.

	Heuristic rule training	Basic training
History display sampling (time on screen in %)	48.5 (18.8)	59.4 (6.0)
Flow meter viewings (samples/min)	0.97 (0.38)	0.83 (0.38)
Control action (interventions/min)	0.68 (0.30)	0.60 (0.33)

Note: Standard deviation is shown in parentheses.

and small performance decrements on secondary tasks. The pattern of results suggests overall that there are advantages associated with HRT but these come at a cost.

There was evidence that HRT led to increased diagnostic accuracy. This confirms the findings of previous research (e.g. Sonntag and Schaper 1988), which has argued that heuristic rules represent effective support for operator activities in fault diagnoses and rectification. The positive effects of HRT were only significant for diagnostic accuracy but not for diagnostic speed. Operators may have adopted accuracy as a high-priority goal because it represented the more important aspect of diagnostic performance and the heuristic rules implicitly encouraged them to focus on it. This could only be achieved at the expense of diagnostic speed, in that the operator only rectified the fault after having carried out very thorough diagnostic checks. Interestingly, the positive effect of HRT on diagnostic accuracy was not limited to novel fault states but also brought benefits to practised faults. This may be indicative of the operational environment placing considerable demands on the operator. The technical training given to both groups did not prepare operators to the extent that they could manage familiar fault states at a high level of reliability, suggesting that even previously practised fault states represented a challenge to participants.

In contrast to the effects on diagnostic accuracy, heuristic rules provided no benefits for system control. There was even cautious evidence for negative effects on this parameter. Two reasons may be put forward to explain this finding. First, the strategies that are needed to be used for fault diagnosis and system control were somewhat opposing. Heuristic rules for fault diagnosis encouraged operators to behave in a reflective manner (e.g. 'Beware confusing one fault state with another!'), whereas those for system control were clearly more action-oriented (e.g. 'Take action even if the situation is unfamiliar!'). These opposing demands may have been difficult to reconcile for operators, resulting in non-optimal utilisation of the heuristic rules. These somewhat conflicting requirements have also been discussed by Landeweerd (1979), whose work revealed that operator abilities in performing these two tasks are independent of each other. Second, the rules for system control may not have been chosen at the same level of specificity, with the heuristic rules for system control being overall more general than the rules for fault diagnosis (e.g. rule 6 vs. rule 1 in Figure 2). Although the rules were derived from a hierarchical task analysis and an error analysis, slight differences in specificity cannot be excluded, which may have contributed to the lack of effectiveness of heuristic rules for system control.

An important experimental finding was the higher demands of HRT on mental resources, which manifested themselves at several levels. First, there were indications of higher fatigue ratings for the HRT-group. This kind of effect may be a sign of increased cognitive demands being placed on the operator (Hockey 1993). Although the fatigue score of HRT was not high in absolute terms, the higher ratings compared to the control group may be indicative of a more active learning mode required to make use of the heuristic rules. Second, there were indications of secondary task performance decrements reaction, with reaction time slowing down and prospective memory failures increasing in the HRT condition. Again, these forms of adaptive responses shown by operators represent a typical indicator of increased cognitive demands for the primary task. These hidden costs of increased task demands are often not detectable in a single task environment but require a multi-task environment to provide a more sensitive test bed (Hockey 1997). The cognitive activities evoked by HRT represented high-level cognitive processing, equivalent to Rasmussen's (1986) knowledge-based system management. The multi-task environment allowed the detection of the potential costs associated with this type of training, which has not yet been reported in previous research applying HRT. This

was because previous work employed a single task or, at best, a dual task environment, often without any clear priorities attached to the task components. The finding of increased cognitive load for the HRT-group indicates the potential costs that this training method incurs. This is an issue that should not be taken lightly, especially as the BT-group did better on all performance measures other than fault diagnosis (albeit the difference was not always significant). This suggests that BT would be less at risk than HRT from a shift in task priorities towards fault diagnosis at the expense of other task elements.

Also of interest were the effects of training on system management strategies. While BT led to a more extensive use of the history display, no such effects were found for flow meter sampling (the data even pointed in the direction of the opposite effect). The history display is a dual function facility (i.e. it represents a monitoring aid to support system stabilisation and also supports fault diagnosis), whereas the flow meter is a single function facility since it has primarily a diagnostic function (i.e. it is not helpful for system stabilisation). The differences in information sampling behaviour reflect the differential effects of training on performance measures discussed above. Previous research has often shown that system management strategies are quite sensitive to training manipulations (e.g. Hockey *et al.* 2007) or, indeed, external conditions such as noise (e.g. Dörner and Pfeifer 1993). The sensitivity of system management strategies corresponds to the predictions by the model of compensatory control (Hockey 1997), suggesting that operators often opt for adaptations in system management to cope with changing operational demands.

Concerning the quality of the operators' explicit mental model (i.e. being able to explain how the technical system works), no difference was found between groups. It may be argued that HRT has a too-strong procedure orientation, which does not lead to improvements of the operator's understanding of the technical system. This may be considered a potential weakness of HRT compared to alternative training methods, which provides better support to operators for adaptive knowledge transfer (i.e. preparing operators for unfamiliar fault scenarios). For example, knowledge-based training (which provides operators with an in-depth understanding of the technical system) was successful in preparing operators for novel fault states on several performance measures (Hockey *et al.* 2007). This increase in performance was paralleled by improvements in the explicit mental model.

Three issues need to be raised to place the findings of this study into an appropriate context: training time; simulation environment; and scope of results.

- (1) There were indications (e.g. unreliable management of familiar fault states) that the study would have benefited from a longer training session. However, extending training time would not have been a feasible option in the current study due to constraints on the availability of the trainee operators, who were given this time off from their vocational training programme. Furthermore, it is difficult to predict the impact of an extension of training time on the findings of the study. Increasing training time does not necessarily lead to performance improvements due to impeding factors such as motivation losses (Hesketh 1997). It is not clear whether a longer training session would have resulted in stronger benefits for HRT compared to BT even if more time had been devoted to the heuristic rules. It also has to be considered that the effective training time of the BT-group was shorter by 30 min because, while the HRT-group was taught the heuristic rules, the BT-group did not engage in performance-relevant activities. If the BT-group had used these 30 min for additional practice, further improvements in performance could have been expected.

- (2) Even though the simulation environment CAMS represents a somewhat artificial setting, the use of such simulation environments (also called scaled worlds) is quite common for modelling complex technical work systems due to the difficulty of carrying out such work in a field setting (Schiflett *et al.* 2004). Although there has been no formal validation of the CAMS environment, the use of a sound theoretical framework for its development (as described in Sauer *et al.* 2000b) is expected to increase the validity of this approach. The findings from other work within the ‘scaled world’ research paradigm have provided evidence for the appropriateness of such simulation environments (Schiflett *et al.* 2004).
- (3) The findings of this research are limited to training contexts in which operators are in the early phases of skill acquisition and have not yet gained much experience of operating the system. The rationale behind HRT is that it should be used with novice operators because expert operators are much less likely to benefit from it as they already tend to make implicit use of these heuristic rules.

Finally, the general utility of HRT is evaluated. Overall, the study provided evidence for some benefits of HRT, though the direct comparison with BT suggested considerable costs associated with the acquisition of the heuristic rules. HRT appeared to be a rather effective method for increasing accuracy in fault diagnosis and rectification, though its utility for system control activities could not be demonstrated. An important finding was the effects of HRT on operator state and secondary task performance, which may be indicative of increased cognitive resource requirements of this training method. This effect may disappear with increasing practice, though concerns may remain that in critical situations (e.g. during the presence of stressors such as noise) where cognitive resources are more limited, this may be a disadvantage (Hockey *et al.* 2007). The potential hidden costs of a training method such as HRT is an issue that has not been raised in previous studies due to their methodological limitations in detecting more subtle forms of changes in operator demands. More work is therefore needed to test the effectiveness of HRT in multiple-task environments.

References

- Amabile, T.M., 1983. The social psychology of creativity. A componential conceptualization. *Journal of Personality and Social Psychology*, 45, 357–376.
- Burkolter, D., *et al.*, 2007. Cognitive requirement analysis to derive training models for controlling complex systems. In: D. de Waard, G.R.J. Hockey, P. Nickel and K.A. Brookhuis, eds. *Human factors issues in complex system performance*. Maastricht: Shaker Publishing, 475–484.
- Burkolter, D., Kluge, A., Sauer, J., Ritzmann, S. and Schüler, K., The predictive qualities of operator characteristics for process control performance: the influence of personality and cognitive variables. *Ergonomics*, under review.
- Collins, A., Brown, J.S., and Newman, S.E., 1989. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In: L.B. Resnick, ed. *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale: Lawrence Erlbaum Associates, 453–494.
- Dörner, D. and Pfeifer, E., 1993. Strategic thinking and stress. *Ergonomics*, 36, 1345–1360.
- Eysenck, M.W. and Keane, M.T., 2005. *Cognitive psychology. A Student's Handbook*. 5th ed. Hove: Psychology Press.
- Frieling, E. and Sonntag, K., 1999. *Lehrbuch Arbeitspsychologie*. Bern: Huber.
- Gopher, D., Weil, M., and Siegel, D., 1989. Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica*, 71, 147–177.
- Hacker, W. and Skell, W., 1993. *Lernen in der Arbeit*. Berlin: Bundesinstitut für Berufsbildungsforschung.

- Hesketh, B., 1997. Lead article dilemmas in training for transfer and retention. *Applied Psychology*, 46, 317–339.
- Hockey, G.R.J., 1993. Cognitive-energetic mechanisms in the management of work demands and occupational 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., Sauer, J., and Wastell, D., 2007. Adaptability of training in simulated process control: comparison of knowledge- and rule-based guidance under task changes and environmental stress. *Human Factors*, 49, 158–174.
- Hollnagel, E., 1998. *Cognitive reliability and error analysis method*. Oxford: Elsevier.
- Kluge, A., et al., 2008. Designing training for process control simulators: a review of empirical findings and current practices. *Theoretical Issues in Ergonomics Science*, in press.
- Landeweerd, J.A., 1979. Internal representation of a process, fault diagnosis and fault correction. *Ergonomics*, 22, 1343–1351.
- Linou, N. and Kontogiannis, T., 2004. The effect of training systemic information on the retention of fault-finding skills in manufacturing industries. *Human Factors and Ergonomics in Manufacturing*, 14, 197–217.
- Lorenzet, S., Salas, E., and Tannenbaum, S.I., 2003. Benefiting from mistakes: The impact of guided errors on learning, performance and self-efficacy. *Human Resource Development Quarterly*, 16, 301–322.
- Morris, N.M. and Rouse, W.B., 1985. Review and evaluation of empirical research in troubleshooting. *Human Factors*, 27, 503–530.
- O'Brien, K.S. and O'Hare, D., 2007. Situational awareness ability and cognitive skills training in a complex real-world task. *Ergonomics*, 50, 1064–1091.
- Patrick, J. and Haines, B., 1988. Training and transfer of faultfinding skill. *Ergonomics*, 31, 193–210.
- Rasmussen, J., 1986. *Information processing and human-machine interaction: an approach to cognitive engineering*. Amsterdam: Elsevier.
- Salas, E., et al., 2006. Design, delivery, and evaluation of training systems. In: G. Salvendy, ed. *Handbook of human factors and ergonomics*. New Jersey: John Wiley, 472–512.
- Sauer, J., Hockey, G.R.J., and Wastell, D.G., 2000a. Effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics*, 43, 2043–2064.
- Sauer, J., Wastell, D., and Hockey, G.R.J., 2000b. A conceptual framework for designing micro-worlds for complex work domains: a case study of the Cabin Air Management System. *Computers in Human Behavior*, 16, 45–58.
- Sauer, J., et al., 2006. Cognitive diversity and team performance in a complex multiple task environment. *Ergonomics*, 49, 934–954.
- Schifflett, S.G., et al., eds. 2004. *Scaled worlds: Development, validation, and applications*. Aldershot: Ashgate Publishing Limited.
- Schoppek, W., 1997. Wissen bei der Steuerung dynamischer Systeme – ein prozessorientierter Forschungsansatz. *Zeitschrift für Psychologie*, 205, 269–295.
- Sonntag, K. and Schaper, N., 1988. Kognitives Training zur Bewältigung steuerungstechnischer Aufgabenstellungen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 32, 128–138.
- Sonntag, K. and Schaper, N., 2006. Wissensorientierte Verfahren der Personalentwicklung. In: H. Schuler, ed. *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe, 261–272.
- Stanley, W.B., et al., 1989. Insight without awareness: On the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology*, 41, 553–577.
- Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.
- Volpert, W., Frommann, R., and Munzert, J., 1984. Die Wirkung allgemeiner heuristischer Regeln im Lernprozess – eine experimentelle Studie. *Zeitschrift für Arbeitswissenschaft*, 34, 235–240.
- Wanous, J.P., Reichers, A.E., and Hudy, M.J., 1997. Overall job satisfaction: how good are single-item measures? *Journal of Applied Psychology*, 82, 247–252.
- Warr, P., 2002. Learning and training. In: P. Warr, ed. *Psychology at work*. London: Penguin, 153–177.
- Wickens, C.D. and Hollands, J.G., 2000. *Engineering psychology and human performance*. Upper Saddle River, NJ: Prentice Hall.