Int. J. Human–Computer Studies (1998) **48,** 729–762

A longitudinal study of the effects of ecological interface design on deep knowledge

Klaus Christoffersen, Christopher N. Hunter & Kim J. Vicente†
Cognitive Engineering Laboratory, Department of Mechanical & Industrial Engineering,
5 King's College Road, University of Toronto, Toronto, Ont., Canada M5S 3G8
e-mail: benfica@mie.utoronto.ca

(Received 26 November 1996 and accepted in revised form 23 February 1998)

Some researchers have argued that providing operators with externalized, graphic representations can lead to a trade-off whereby deep knowledge is sacrificed for cognitive economy and performance. This article provides an initial empirical investigation of this hypothesis by presenting a longitudinal study of the effect of ecological interface design (EID), a framework for designing interfaces for complex industrial systems, on subjects' deep knowledge. The experiment continuously observed the quasi-daily performance of the subjects' over a period of six months. The research was conducted in the context of DURESS II, a real-time, interactive thermal-hydraulic process control simulation that was designed to be representative of industrial systems. The performance of two interfaces was compared, an EID interface based on physical and functional (P + F) system representations and a more traditional interface based solely on a physical (P) representation. Subjects were required to perform several control tasks, including startup, tuning, shutdown and fault management. Occasionally, a set of knowledge elicitation tests was administered to assess the evolution of subjects' deep knowledge of DURESS II. The results suggest that EID can lead to a functionally organized knowledge base as well as superior performance, but only if subjects actively reflect on the feedback they get from the interface. In contrast, if subjects adopt a surface approach to learning, then EID can lead to a shallow knowledge base and poor performance, although no worse than that observed with a traditional interface.

1. Introduction

The purpose of this article is to investigate how operators' deep knowledge is affected by ecological interface design (EID), a theoretical framework for human–computer interface design for complex human–machine systems (Vicente & Rasmussen, 1990, 1992; Vicente, Christoffersen & Hunter, 1996). EID has a family resemblance to other approaches to interface design (e.g. Coekin, 1969; Goodstein, 1981; Woods, 1991; Bennett & Flach, 1992) in that one of its goal is to exploit the power of human perception by relying on graphical representations. The basic idea is that operators should be able to satisfy task demands more effectively and with less effort if they can deploy their powerful perceptual systems rather than their more limited cognitive resources. Several empirical evaluations of EID have been conducted to date in a number of diverse domains, including process

---

†Author to whom correspondence should be addressed.

1071-5819/98/060729 + 34$25.00/0/hc980190

control (e.g. Vicente, Christoffersen & Pereklita, 1995; Pawlak & Vicente, 1996; Christoffersen, Hunter & Vicente, 1996, 1997; Janzen & Vicente, in press), neonatal intensive care units (Sharp, 1996), and hypertext information search (Xu, 1996). These studies have shown that interfaces based on the principles of EID can lead to better performance than more traditional interfaces, both under normal and abnormal conditions.

Nevertheless, the family of approaches of which EID is an exemplar is not without its potential drawbacks. Wickens (1992), for instance, has argued that "those features of an interface that may reduce effort and increase performance may actually *reduce* [long-term] retention" (p. 843, emphasis added). The rationale for this claim is as follows. If an interface is intended to reduce operator effort by displaying a graphical representation of the process, operators are not forced to extensively study, and thus completely understand, plant functionality since it is already apparent in the surface features of the interface. In contrast, interfaces that do not present this information at the surface require greater cognitive effort to derive and comprehend system functionality and relationships, thereby inducing learning of deep structures. Therefore, interfaces such as those based on the EID framework may indeed require less effort and lead to better performance, but they may do so at an important cost—a reduction in the acquisition of deep knowledge. How plausible is this argument?

On the one hand, one could argue that Wickens' (1992) claim cannot possibly be correct. After all, in the extreme, it would suggest that we should deliberately design impoverished interfaces to force operators to think, thereby inducing the acquisition of deep knowledge. A somewhat similar philosophy was voiced by a training director interviewed by Seminara, Gonzalez and Parsons' (1977) in their seminal study of human factors in nuclear power plant control rooms: "The [operator's] job is boring so the board creates a challenge. If the board is too straight-forward or 'human engineered', then the operator will lose his edge" (p. 21-1). Clearly, this is a radical view that contradicts the very spirit of cognitive engineering.

On the other hand, it is important to point out that Wickens (1992) is not alone in his opinion. Similar concerns regarding the potential pedagogical drawbacks of external representations have been voiced in the education literature (e.g. Pea, 1993; Salomon, 1993). The argument put forth there is similar to Wickens'. Because external representations are intended to off-load cognitive burden from the student to artifacts in the environment, they require less of the student. As a result, cognition unaided by tools may lead to deeper understanding than cognition aided by external representations.

Even more important, however, is the fact that Wickens (1992) cited the results of approximately half a dozen empirical studies to support his claim. In fact, one of those studies was the first empirical evaluation of the EID framework (Vicente, 1991). Thus, his claims, although perhaps counterintuitive to some, cannot be so easily dismissed. On the contrary, they deserve systematic empirical scrutiny.

It is important to note that the studies cited by Wickens (1992) were not originally designed to explore the potential trade-off between effort and performance on the one hand, and deep knowledge on the other. For example, all of the studies were conducted over a relatively short period of time. However, it is possible—perhaps even likely – that changes in subjects' deep knowledge would only occur over longer periods of time. If so, then degradation or enhancement of deep knowledge would not be reliably observed in studies of relatively brief durations. In addition, the experiments cited by Wickens also tended to examine the aggregate performance and learning of groups of subjects. It is possible that the effects of interface design on deep knowledge are individual-specific. Subjects with different background knowledge or attitudes to learning may react differently to the same interface. Some subjects may exhibit the trade-off posited by Wickens, but others may not. Such individual differences would be masked in an aggregate analysis. Thus, we believe that the empirical support for Wickens' claim is, at best, indirect.

In this article, we describe an empirical investigation that was designed to overcome these limitations. The purpose of the study was to investigate the long-term impact of EID on deep knowledge. Two interfaces for a thermal-hydraulic process control simulation were compared, one based on EID and another based on a more traditional interface format. Subjects were required to control the simulation under both normal and abnormal conditions with one of the two interfaces. The experiment was conducted on a continuous, quasi-daily basis for a period of six months so that slow changes in subjects' deep knowledge could be observed. In addition, both performance and deep knowledge were analysed on a subject-by-subject basis so that any pertinent individual differences could also be observed.

The following section provides the background for the experiment, including descriptions of the EID framework, the simulation that was used as a research vehicle, and the two interfaces that were evaluated.

2. Background

2.1. THEORY

The theoretical foundations of the EID framework have already been described at great length elsewhere (Vicente & Rasmussen, 1990, 1992; Vicente *et al*., 1996), so only a brief summary will be provided here. The goals of EID are twofold: to allow people to deal with the demands of complex work domains by relying on their powerful perception–action systems whenever possible, and to simultaneously provide the support people require to engage in the more effortful analytical problem-solving activities that are needed to adapt to unanticipated contingencies (e.g. faults). There are two design phases in operationalizing this goal into a particular interface, the first dealing with the identification of the information that needs to be represented in the interface, and the second dealing with the visualization of this information and the means for control.

The first phase of EID is perhaps the most unique, since it adopts the abstraction hierarchy (AH) formalism developed by Rasmussen (1985) to develop a representation of the work domain. Each level in the hierarchy is a different model describing the goal-relevant constraints of the same work domain. Higher levels of abstraction represent the work domain in functional terms and lower levels represent the work domain in physical terms (Vicente & Rasmussen, 1992). For example, for process control systems, five levels have been found to be of use (Rasmussen, 1985) : *functional purpose*, describing the overarching purposes for which the work domain was designed; *abstract function*, describing the first principles of the work domain in terms of mass and energy conservation laws; *generalized function*, describing the functions that have been built into the work domain; *physical function*, describing the components and equipment in the work domain; and *physical form*, describing the location and appearance of that equipment.

One of the most important features of the AH is that adjacent levels are connected by means–ends links [see Bisantz and Vicente (1994) for a detailed example for the testbed used in this study]. As a result, the level above any other specifies *why* a particular component or function exists in the structure of the work domain (i.e. what ends it can fulfill), and the level below any other specifies *how* a particular function has been structurally implemented (i.e. what means are available to satisfy it). In other words, the linkages between levels in the hierarchy show the relationship between components (what can be acted upon) and purposes (what needs to be achieved). As we will discuss shortly, this representation of functional structure has properties that seem desirable from the viewpoint of helping operators develop deep knowledge of the work domain.

The second phase of EID is more common. Once an AH for a particular work domain has been developed, EID suggests that the abstract relationships identified in this hierarchical problem representation be mapped onto concrete visualizations that reveal these constraints in a manner that can be easily and efficiently picked up by human perceptual systems. Thus, EID provides a set of normative device models, not via training (as in Kieras & Bovair, 1984) but via an external graphical representation, that can be used to learn to control a complex system. In addition, operators should be able to directly act on this representation in a direct manipulation fashion. The idea, then, is to leverage human perception–action skills to develop a faithful, transparent representation of the functional and physical characteristics of the work domain.

What is the intended impact of these recommendations on operators' deep knowledge? EID makes multiple models of the work domain, as well as the relationships between those models, visible in the interface. If attended to, this feedback provides a strong basis for the development of deep knowledge. Because operators can directly observe the functional structure of the work domain in the interface, they may be able to develop a functionally organized knowledge base with experience. Note that this rationale contradicts Wickens' (1992) arguments. Instead of expecting a deficit in deep knowledge caused by an externalized graphical representation, we would instead expect an enhancement of deep knowledge.

Later in this section, an example of an interface based on the principles of EID will be presented. First, however, the simulation that has been adopted for this research will be described.

2.2. DURESS II

DURESS (DUal REservoir System Simulation) II is a real-time, interactive thermal-hydraulic process control microworld that we have been using in our laboratory as a testbed to evaluate the EID framework. DURESS II was designed to be *representative* (Brunswik, 1956) of complex work domains, thereby promoting generalizability of results to industrial systems (Vicente, 1991).

The physical structure of DURESS II is illustrated in Figure 1. The plant consists of two redundant feedwater streams (FWSs) that can be configured to supply water to either, both, or neither of the two reservoirs. Each reservoir has associated with it an

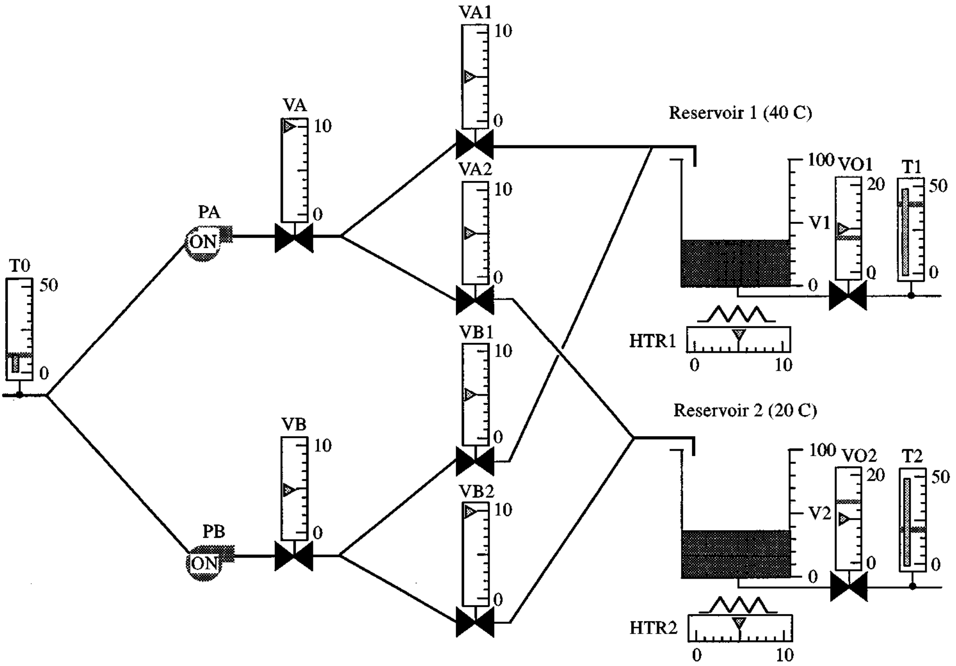

FIGURE 1. P interface for DURESS II.

externally determined demand for water that can change over time. The system purposes are two-fold: to keep each of the reservoirs at a prescribed temperature (40 and 20°C), and to satisfy the current mass (water) output demand rates. To accomplish these goals, the subject has control over eight valves (VA, VA1, VA2, VO1, VB, VB1, VB2 and VO2), two pumps (PA and PB) and two heaters (HTR1 and HTR2). All of these components are governed by first-order lag dynamics, with a time constant of 15 s for the heaters and 5 s for the remaining components. The system input temperature ($T0$), reservoir output temperatures ($T1$ and $T2$), and the volumes for both reservoirs ($V1$ and $V2$) are also displayed in Figure 1.

2.3. P INTERFACE

Figure 1 is one of the interfaces used in this study. It provides a physical (P) representation of DURESS II, displaying only the state of the physical components and the goal variables. The first meter on the extreme left of the display is a thermometer ($T0$) measuring the inlet water temperature. The vertical bar increases in height as the water temperature increases. The normal inlet water temperature is 10°C, as indicated by the thin area on the $T0$ scale. After the thermometer, the input water stream splits and flows to two pumps (PA and PB) that operate as discrete switches (on or off). The subject uses a mouse to click on the pump to change its state. The pumps are displayed in black (with white lettering) if they are off, and in light gray (with black lettering) if they are on. The maximum flow rate through each pump is 10 units/s. If either pump is turned on without any of the downstream valves being opened, the pump will fail after approximately 5 s. This error terminates the trial.

The next set of components are the primary valves (VA and VB), having a continuous range of 0–10. The valve state is set using a mouse to either drag the triangular pointer to the desired setting, or to simply click on the scale at the desired point. From these primary valves, each FWS splits into two secondary valves connecting each stream to both reservoirs. The secondary valves (VA1, VA2, VB1 and VB2) operate in the same manner as the primary valves. The water then flows to each of the two reservoirs, where it is heated and removed, through the use of the heaters (HTR1 and HTR2) and the output valves (VO1 and VO2), in order to meet the temperature and demand goals, respectively. The reservoirs have a maximum capacity of 100 units. Reservoir volume levels are indicated by a scale on the side of each reservoir and by the shaded area depicting water in the reservoir. It is possible to overflow either of the reservoirs, if input flow rate is consistently greater than output flow rate. When reservoir volume exceeds the maximum capacity of 100 units, the trial ends automatically.

The heaters (HTR1 and HTR2) also have a continuous range of 0–10. The subject can either slide the triangular pointer on the heater scale to the desired setpoint, or click on the scale itself at the desired point. Heating an empty reservoir for an extended period will lead to a malfunction. Thus, if there is continued heat transfer to a reservoir without any water in that reservoir, then the system will fail and the trial will end. The water temperature in the reservoirs is displayed with thermometers ($T1$ and $T2$). The goal temperature is represented as a thin area on the temperature scale. There is a tolerance of $\pm$ 2°C from the setpoints (40°C for Reservoir 1 and 20°C for Reservoir 2). If the water in the reservoir boils, the system fails and the trial ends.

Finally, the subjects also have control over the outlet valves (VO1 and VO2) that are used to meet the demand goals. These valves operate in the same manner as the other valves, except that their maximum setting is 20. The demand for each reservoir is indicated by a thin area on the valve setting scale. This goal area, which is $\pm$ 1 unit around the desired level, moves as a function of changes in the demand.

This interface design format was chosen as a baseline condition for our experiment because it is typical of existing interfaces for process control systems (e.g. Malone, Kirkpatrick, Mallory, Eike, Johnson & Walker, 1980; Becker, 1991; Vicente, Burns, Mumaw & Roth, 1996).

2.4. P + F INTERFACE

The P + F interface, based on the principles of EID, is illustrated in Figure 2. A very detailed graphical description can be found in Pawlak and Vicente (1996), so only a brief overview will be provided here. The input water thermometer, both pumps, and all of the valves operate in the same manner as in the P interface. However, the P + F interface also contains higher-order functional information identified through an AH analysis of DURESS (see Vicente & Rasmussen, 1990) that describes the state of the functions that the physical components are intended to achieve. Thus, each valve also has a flow meter next to it (FVA, FVB, FA1, FA2, FB1, FB2 and MO1 and MO2 for the mass output flow rates). These meters have the same value range as their respective values.

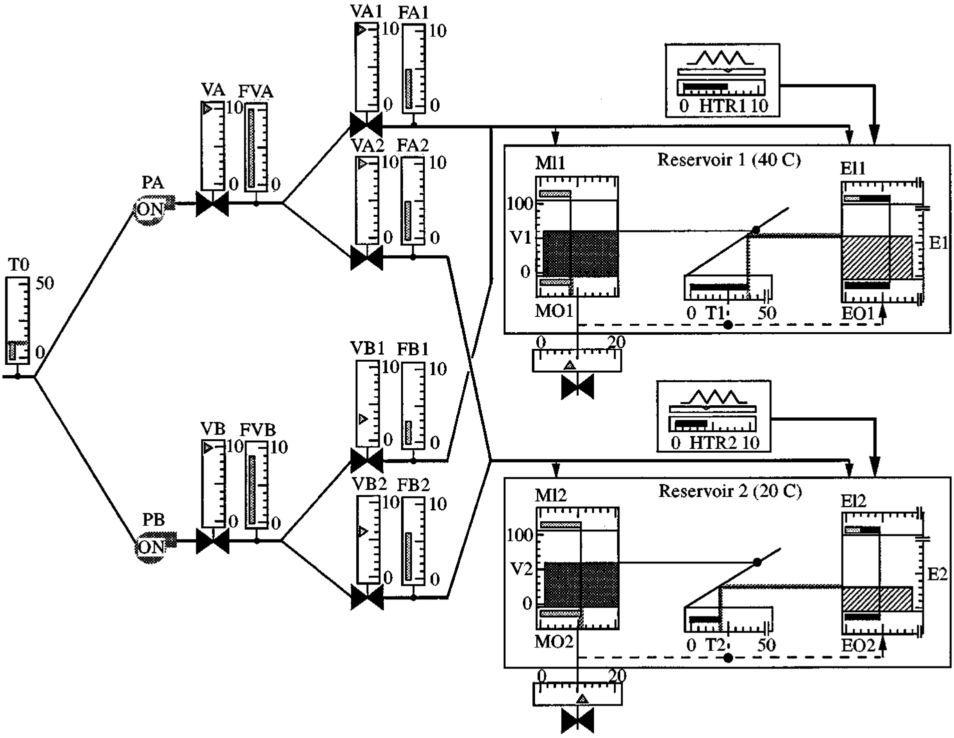

FIGURE 2. P + F interface for DURESS II.

The boxed group of graphics on the right of Figure 2 provides additional higher-order functional information in the form of first principles (i.e. mass and energy conservation laws). The rectangular graphic on the left represents the mass balance (i.e input flow rate, inventory and output flow rate) for the reservoir and the graphic on the right represents the energy balance. Both representations operate in a similar manner. Referring to Reservoir 1, the various inputs are shown at the top of the graphics (MI1 for mass and EI1 for energy). Inventories for each representation are indicated by scales on the side of each graphic (V1 for volume/mass and E1 for energy). The outputs, MO1 for mass and EO1 for energy, are shown at the bottom of each graphic. The energy inputs to each reservoir (EI1) are partialled out according to the two contributors. Thus, the energy added by the FWS is shown as the lightly shaded bar, and the energy added by the heater is shown as a dark bar. The mass and energy graphics rely on a funnel metaphor that is created by the line connecting the input and output. The slope of the line represents the rate at which the mass (or energy) inventory should be changing. For example, if input equals output, as is the case for the mass and energy balances for both reservoirs in Figure 2, then the line is perpendicular, indicating that the level should not be changing. If the bottom is wider than the top (i.e output > input), then it should be easy to visualize the consequence: viz., that volume should decrease.

The graphic in the middle, between the mass and energy balances, illustrates the relationship between mass, energy and temperature. There is a horizontal line with a ball on the end that emanates from the current mass inventory level ($V1$). Changes in the height of this line always accompany any change in mass inventory (i.e. the bar will always be at the same height as the water level, ($V1$). The diagonal line in the centre display rotates about its leftmost endpoint (connected to the top left of the $T1$ box) and is always tangent to the ball on the end of the horizontal line. Thus, a change in the vertical position of the horizontal line serves to change the slope of the diagonal line in the centre display. For example, if volume increases, the horizontal line goes up, causing the diagonal to rotate counterclockwise, increasing the slope of the diagonal line. The slope of the diagonal line represents the function that maps the relationship between mass and energy onto temperature. This mapping is indicated by the line emanating from the current energy inventory level ($E1$) that comes across and reflects off of the diagonal line at a right angle down onto the temperature scale ($T1$). The goal temperature is indicated by the thin shaded area on the temperature scale. This goal area reflects back from the temperature scale, off of the diagonal line, and onto the energy inventory scale. In addition, off-scale markers are added to the output temperature scales and the energy input, inventory and output scales as well. These were added to the interface by creating a gap in the scale at the off-scale point, thereby allowing subjects to discriminate the maximum value from off-scale (Mumaw, Woods & Eastman, 1992).

2.5. RELATIVE IMPACT ON DEEP KNOWLEDGE

How will the P and P + F interfaces affect subjects' deep knowledge? A strong interpretation of Wickens' (1992) argument would predict that subjects with the P + F interface would develop an impoverished (or even non-existent) deep understanding of DURESS II because they would have become accustomed to using the surface features as a crutch. As a result, their deep knowledge would actually be *inferior* to that of subjects using the P interface. In contrast, the EID framework would predict that the P + F subjects who attend to the information in their interface should demonstrate a deeper knowledge of DURESS II than the P subjects because the former are given a more comprehensive visualization of the set of goal-relevant constraints that govern the process. The experiment described next is an initial empirical investigation of these hypotheses.

3. Method

3.1. EXPERIMENTAL DESIGN

A repeated measure, between-subjects design, with interface (P or P + F) as the independent variable, was adopted for this experiment.† Subjects were assigned to one of the two

---

†It might be argued that there is a confound in this experimental design because the P + F interface has more information than the P interface (cf. Maddox, 1996). This interpretation is incorrect because one of the defining features of the EID framework is that it helps identify what information needs to be displayed. As a result, information content is not a factor that should be held constant, but rather a manipulation whose effects we are interested in investigating (cf. Vicente *et al*., 1996).

interfaces and participated for a total of six months. At the end of the experiment, a transfer manipulation was conducted, with subjects controlling the system for six trials with the alternate interface.

3.2. SUBJECTS

Six male subjects, ranging in age from 23 to 32 years participated in this study. The subjects, with one exception, had science or engineering backgrounds. An attempt was made to assign subjects to interface groups so as to match for background. Each subject was paid $5 per session. These regular "wages" were paid roughly every six weeks. A bonus of $2 per session was offered for completing the entire experiment. Extra bonuses were also offered for "good" performance, although subjects were not given any details about how performance would be measured or evaluated. These additional bonuses were only paid upon completion of the experiment, and were designed to maintain subjects' motivation over the course of the experiment. In the end, the extra bonuses were divided equally among the subjects.

3.3. APPARATUS

The DURESS II simulation runs on a Silicon Graphics IRIS Indigo R4000 computer workstation. The simulation code was written in C, while the two interfaces were constructed using a graphical construction set called FORMS. All data collection and analyses were performed using this computer, as well as a Silicon Graphics IRIS Indigo R4400, an Apple Macintosh IIvx computer, and a pair of Apple Macintosh Quadra 660AV computers.

3.4. EXPERIMENTAL TASKS

Subjects were asked to perform four different types of control tasks; startup, tuning to new output demand setpoints, shutdown and fault management [see Christoffersen *et al.*, (1996, 1997) for a detailed description]. For all control tasks but shutdown, subjects had to achieve steady state which was defined as maintaining both reservoirs in the goal areas (both temperature and output demand) for five consecutive minutes. If subjects performed actions that harmed the system (e.g. spilled water from the reservoirs, let the water in the reservoirs boil, heated an empty reservoir for an extended period or had a pump on for more than a few seconds with no open downstream path available), then the trial would end prematurely and an error message describing the problem was displayed. There were two types of faults. Routine faults occurred nine times during the experiment, and were intended to represent comparatively simple recurring failures in an industrial process. Non-routine faults only occurred towards the end of the experiment, and were intended to be more complex, multiple faults that are equivalent to rare events in industrial systems [see Christoffersen *et al.* (1997) for a description of each].

3.5. PROCEDURE

There were four distinct phases in the experiment: an introductory session where the experimental protocol was explained to subjects, an introductory practice phase (one month) during which the complexity of tasks was gradually increased, an extended practice phase (four months) characterized by repeated exposure to normal trials and occasional exposure to routine faults and a final phase (one month) examining the long-term effects of each interface. The latter phase included three non-routine faults and the six transfer trials as well (see below). The entire experiment lasted approximately six months, during which subjects came in roughly every weekday for about an hour, performing a total of 224 trials. Six test days were also set aside during the experiment to have the subjects perform a set of knowledge elicitation tests. The experiment concluded with a debriefing session, attended by all but one of the subjects.

The knowledge elicitation tests might have encouraged some subjects to reflect about their performance, so the tests were granted their own sessions (referred to as test days). Our goal was to minimize the influence of previous trials on the test results and of the writing of the tests on the results of the subsequent trials. Furthermore, the sessions were relatively infrequent so as not to induce subjects into artificially thinking more about the work domain than they would otherwise. The way in which the test days fit into the context of the experiment is summarized in Table 1.

3.6. DEPENDENT VARIABLES

Many different dependent variables were collected during the study. For the sake of clarity, only those whose results are reported later will be described [see Christoffersen, Hunter and Vicente (1994) for a complete description].

Informal comments from subjects were recorded throughout the experiment. Although highly subjective, these comments provided us with a valuable source of insights regarding subjects' perceived level of motivation and the criteria they seemed to adopt in performing the required tasks. These data thereby served as a source of hypotheses about individual differences that could be confirmed or disconfirmed by analyses of more objective dependent variables.

Various measures of performance on both normal and fault trials were collected and have already been reported in other articles based on this experiment (see Christoffersen *et al*., 1996, 1997). In this article, we report the means and standard deviations of completion times for an early and late block of normal trials, as well as the means and standard deviations of completion times for the routine and non-routine fault trials.

TABLE 1

Times at which test days occurred, as a function of the number of trials of control experience and the number of faults experienced

Test No.	No. of Trials	Cumulative	No. of Faults	Cumulative
1	0	0	0	0
2	21	21	0	0
3	46	67	1	1
4	83	150	4	5
5	67	217	7	12
6	6	223	2	14

These data provide a brief summary of the comparative skill level of the subjects, thereby serving as a context for interpreting the results from the knowledge elicitation measures.

A between-interface transfer manipulation was conducted at the end of the experiment, thereby giving subjects four normal trials and then two fault trials with the other group's interface. Interface transfer effects are an indirect measure of deep knowledge (Kossack, 1992). Subjects who can transfer their skills across interfaces probably have a deep understanding of the process that is independent of the interface they had been using. In contrast, subjects who cannot transfer their skills across interfaces probably have a shallow understanding of the process that is dependent on the surface features of the interface they had been using.

The other knowledge elicitation measure reported in this article is known as a control "recipe" (Irmer & Reason, 1991). Subjects were asked to write down a set of instructions describing how they control the system. These instructions were to be sufficiently detailed to allow someone who has never seen the system before to control it in the same manner as the subjects. Only control recipes for doing a startup task were solicited. The recipe test was given six times, as a part of each test day (see Table 1). The content and form of the recipes were expected to give insight into: the extent to which each subject acquired knowledge over time; the extent to which each subject became sensitive to specific features of the process; and the extent to which their knowledge was organized in a deep or shallow manner.

4. Results

This section is divided into four subsections corresponding to results from: subjects' subjective comments, performance on normal and abnormal trials, transfer trials and control recipes. As already mentioned, the performance results had already been reported in previous articles based on this experiment (Christoffersen *et al.*, 1996, 1997). All of the other analyses presented in this section are reported for the first time.

The analysis of the results is unorthodox and deserves explanation. As with other related studies that we have conducted (e.g. Pawlak & Vicente, 1996), we tried to apply the principle of *representative design* (Brunswik, 1956) in this experiment to improve the generalizability of results to operational settings. This choice greatly increased the complexity and duration of the experiment. As a result, only a very small number of subjects could be included in each group in the time available. Furthermore, we suspected that the impact of the interfaces on deep knowledge might depend on each subject's level of motivation and idiosyncratic background knowledge (see Section 1). For these reasons, we analysed the results for each individual subject separately and in detail rather than conducting statistical tests on group averages.

4.1. SUBJECT PROFILES

A concerted effort was made to record any comments that subjects made throughout the experiment. These data allowed us to develop a comparative subjective assessment of subjects' level of motivation and interest in the experiment. In a study of this duration, differences in subjects' willingness to learn and devote effort to reflecting on the task become clear over time. Previous research has shown that students' learning styles are significantly correlated with learning outcomes (Moore, Smith & Telfer, 1994). In the absence of a formal measure, the experimenters' subjective assessments of each subject's level of learning motivation will be presented (only within-group rankings will be presented, since there was only a clear consensus on these). This assessment is admittedly subjective and should therefore be viewed as a source of hypotheses that can be confirmed or disconfirmed by other analyses described later. The discussion is organized by interface group, beginning with the P group.

4.1.1. P group

The subject profiles for the P group will be presented according to the perceived level of motivation, with the subject judged to be the most highly motivated being discussed first.

TL was by far the most competitive subject. He often said that he dreamt of new methods in order to reduce his trial completion times. After he got his times as low as he thought possible, he began to minimize the number of control actions that he used. He frequently modified or tried a new method to improve his performance, which may have resulted in a greater than average number of instances where he accidentally caused the system to fail. At trial 200, he was still trying out new methods to improve his performance. Once, he even brought coffee to be used as a reward for good performance; he would not let himself have a drink until he reached steady state!

WL tried hard to achieve goal states for the system. He considered time important but not vital. He often was not forthcoming with what he was thinking and became frustrated when he was asked; he had trouble verbalizing his rationale for action. He consistently said that he had a lot of trouble understanding the heaters. Sometimes he focused his attention solely on one part of the system for an extended period, ignoring the rest. This usually led to poor performance.

ML was one of the two subjects who had a very limited physics background. In addition to this limited background, he also appeared to put forth the least effort of all the subjects in the experiment. During many of the trials, he read a magazine or focused on anything but the simulation at hand, while impatiently waiting for the system to reach the goal conditions, or for the demand setpoints to change. Throughout the course of the experiment, it was very difficult to get him to come in on time. In fact, there were large stretches of time during which he completed few trials. As a result, ML completed the experiment well after the other subjects.

Despite his apparent lack of effort and interest in the experiment, he managed to have some of the fastest shutdown times of all subjects. Part of this may be due to his experimentation with the system. He sometimes blew up the system just to see what it look, intentionally testing its limits. However, towards the end of the experiment, he often accidentally blew up the system by heating an empty reservoir (a sign of draining the reservoirs prematurely during shutdown).

4.1.2. P + F group

The subject profiles for the P + F group will also be presented according to the perceived level of motivation, with the subject judged to be the most highly motivated being discussed first.

IS was the other subject with very little physics background (his previous schooling was primarily in biology). He tried to keep as clear a mind as possible when doing the experiment, so that he could react to all task conditions without confounding them with his self-proclaimed weak understanding of thermal-hydraulics. Thus, he did not assume that the system followed the laws of physics. He learned about the system by observing the graphics in order to achieve the goals. As the experiment progressed, he formed rules that he could use to control the system by experimenting with various system parameters (e.g. "What happens if I overflow the reservoir?"). However, he did not use the rules until he was convinced that the parameters that the rules depended on would remain stable throughout the experiment. Therefore, it is likely that IS' progress was slower than it would have been had he used his new rules when he first discovered them. Essentially, he was methodical and cautious in his approach to learning about the system. IS seemed to be very focused during the study. During one fault, his hand became so tense that he had to shake it out. Like most subjects, he tried to minimize time to some extent. However, he claimed never to sacrifice the quality of his performance just to be able to leave early.

AV would experiment with the system in order to find its limits. For example, he would empty a full reservoir just to see how long it look, even though this meant that the trial took longer. His justification for exploration was to ultimately spend as little time in the experiment as possible; time was his prime criterion for performance. As a result, if AV believed he had gone outside of the goal boundaries while trying to maintain steady state (meaning the trial would go at least an extra 5 min), then he would appear to purposely bring about a system failure (see Section 3). After the experiment, he confirmed that he did this. This is not surprising since he complained the most out of all of the subjects about how boring the experiment was.

AS seemed to control the system simply by observing the graphics describing mass and energy flows. He said that he did not really think about what the system was doing, as long as these graphics were indicating that the system goals were being met. In fact, he ignored the valves and pumps on the left-hand side of the display after startup, failing to consider that this strategy could degrade performance during faults. Though he often expressed ideas that he tried to use to speed up his performance, it did not seem as though he pursued them very actively. He seemed more interested in finishing the trials quickly. However, at the beginning of the experiment, he noted that he would "play around" with the system in order to see how it responded. This often resulted in slower times. In general, AS put forth a nominal level of effort towards the completion of the experiment.

4.1.3. Discussion

By chance, the two subject groups seemed to be roughly balanced in terms of level of motivation. TL and IS took the experiment very seriously and made a consistent effort to learn and improve their performance (although clearly, TL was the most highly motivated of the lot). AS, and especially ML, treated the experiment as a chore and did not seem to reflect very much on their performance. WL and AV exhibited an intermediate level of motivation, showing some signs of creativity and insight but not having the consistent devotion exhibited by TL and IS.

4.2. PERFORMANCE DIFFERENCES

Although the focus of this paper is the effect of interface on subjects' deep knowledge, it is necessary to provide a very brief summary of the relative performance differences between individuals on normal and fault trials. These results provide a more objective evaluation than the subjective assessments just presented, as well as providing a context for interpreting the results from the knowledge elicitation measures described in the remainder of the article.

Table 2(a) shows the results for the first 22 and last 20 normal trials, not including the transfer trials.† The means and standard deviations for the time to complete startup tasks are indicated. Beginning with within-group comparisons, AV and TL are clearly the fastest in their respective groups. AS is clearly the slowest in the P + F group. ML was the slowest in the P group early on, whereas WL was the slowest at the end of the experiment. As for between-group comparisons, the following pairs of subjects demonstrated very similar absolute levels of mean performance: AV and TL, IS and ML and AS and WL. These subject pairs can be considered cohorts in the sense that they exhibit a comparable level of performance with different interfaces after extensive practice. Note, however, that the P + F subjects demonstrated a high level of consistency. This variability effect was statistically significant and was observed throughout the experiment, up to the transfer trials at which point the P subjects, who were now using the P + F interface, became less variable than the P + F subjects who were now using the P interface (Christoffersen *et al*., 1996).

Table 2(b) shows the results for the routine and non-routine faults, again measured by time to complete the trial. These data are a better indication of skill in that fault trials, especially the non-routine type, were more challenging than the normal trials. Here we find that AV was again the fastest in the P + F group. TL was the fastest in the P group for routine faults, but WL was the fastest for non-routine faults. AS and ML were the slowest again in their respective groups. In contrast to the normal trials, the fastest subjects in each group are no longer comparable; AV was by far the fastest overall. In fact, if one matches the subjects across groups as cohorts in terms of performance, one finds that the P + F subject was always faster than the respective P subject. This pattern held even for the slowest pair of subjects; AS was faster than ML, who actually was only able to complete one of the three non-routine faults.

4.2.1. Discussion

These performance differences are generally in line with the subject profiles in the previous subsection, with one exception. The two subjects who were perceived to be the least motivated, AS and ML, were the slowest in their respective groups. TL, who was perceived to be the most motivated of all of the subjects, was usually the fastest in the P group, although this pattern was more prominent for normal faults. The only strong discrepancy was the fact that IS, who was perceived to be the most motivated of the P + F subjects, was not the fastest in his group. Instead, AV was consistently the fastest and, on the challenging fault trials in particular, was head and shoulders above all of the

---

[†] Most subjects did not successfully complete several trials at the beginning of the experiment. Thus, a larger number of trials was included in the first block to achieve a balanced comparison.

TABLE 2
Summary of performance differences on (a) normal trials and (b) fault trials

(a) Average startup task completion time for first 22 and last 20 normal trials (from Howie, 1996)

		Trials 1–22		Trials 196–217	
Group	Subject	Mean	S.D.	Mean	S.D.
	AS	860.2	441.3	437.2	31.3
P + F	AV	517.3	149.0	353.5	16.1
	IS	644.4	125.8	399.0	17.5
	ML	660.3	261.8	390.2	48.0
P	TL	493.6	80.2	357.5	20.5
	WL	624.3	170.4	437.2	97.9

(b) Average trial completion time for routine and non-routine faults

		Routine		Non-Routine	
Group	Subject	Mean	S.D.	Mean	S.D.
	AS	868.3	362.6	1166.0	164.0
P + F	AV	570.0	127.3	740.3	89.3
	IS	667.0	140.8	1030.5	101.1
	ML	883.3	314.4	1255.0	†
P	TL	686.5	264.3	1172.0	93.3
	WL	687.8	164.6	1022.0	53.7

†This subject was only able to successfully complete 1 of the 3 non-routine fault trials.

other subjects. Perhaps, it was AV's deliberate attempts at exploratory behaviour that led him to achieve such a high level of performance.

If we compare the subjects by cohorts, the P + F group seems to have a slight advantage on fault trials. The fault detection, diagnosis and compensation results show a stronger pattern in favour of the P + F group (Christoffersen *et al*., 1997), thereby reinforcing the advantage observed on normal trials. In the analyses described next, we will see if this performance advantage is accompanied by a deficit in deep knowledge.

4.3. TRANSFER TRIALS

With two exceptions, the results from the transfer trials are not very informative and so only a brief summary will be provided here. The general pattern seemed to be that experience with the P + F interface can help some subjects overcome the deficiencies of the P interface to some extent. On the other hand, experience with the P interface did not seem to provide much of an advantage for transfer performance. Furthermore, no subject improved upon transferring to the P interface, whereas transferring to the P + F interface led to improved performance for some.

The most important results from the transfer trials were obtained with AS and ML. When AS moved from the P + F to the P interface, he had a great deal of difficulty completing the task on the first transfer trial, and on subsequent trials was not able to stabilize the system at all, reaching the 45 min time limit set for trial completion. One interpretation of this result is that AS used the surface features of the P + F interface as a crutch, and as a result, when he moved to the P interface, he was no longer able to do the task because his crutch was gone. This explanation is consistent with the observation that AS did not put a great deal of effort into the experiment and was consistently the least proficient performer in the P + F group. Further evidence in favour of this interpretation will be provided later. When ML moved from the P to the P + F interface, he went from being the worst in his group to far and away the best, especially on the two fault transfer trials. While the other two P subjects had trouble adapting to the new interface, ML used it very effectively and efficiently.

4.4. CONTROL RECIPES

The control recipes that subjects were occasionally asked to write out (e.g. Tables 6–11) provided a great deal of insight into the extent and organization of their knowledge. The way in which the recipes are written, and especially how their form changes over time, can provide important clues as to the way in which subjects think about the process. The degree of chunking/enrichment measures, at a coarse level, the extent to which subjects acquired knowledge over time. The degree of differentiation measures, at a more detailed level, the extent to which subjects became sensitive to specific features of the process. Finally, the organization of knowledge directly measures the extent to which subjects' knowledge was organized in a deep or shallow fashion.

4.4.1. Chunking/enrichment

Before we try to determine if subjects' knowledge was deep or shallow, it is useful to investigate the acquisition of knowledge during the experiment. The length and grouping of the control recipes were used as very coarse measures of the degree of chunking or enrichment in knowledge over time. These results are presented in Table 3.

Table 3(a) shows the length of the recipes, measured by the number of words. The most salient finding is that the length of the control recipes does not vary monotonically over time. Instead, for each subject, there seems to be an oscillation between expansion and reduction in length. It is possible that expansion is indicative of enrichment, i.e. acquiring new, more detailed knowledge. Similarly, length reduction could be interpreted as abstraction, i.e. the aggregation of information into a simpler form. Both of these patterns can be observed if one examines the content of the recipes. The acquisition of new knowledge is expected and is documented in more detail below. Abstraction was also observed in various forms. For example, instead of listing all the valves individually, with experience some subjects would collectively refer to "the valves".

We can also examine how subjects chunked their recipes. In all cases but one, subjects explicitly organized their recipes into a distinct number of chunks, either by labeling each step numerically or by distinguishing it typographically (e.g. by a dash or bullet). Table 3b shows the results of this analysis. Again, we find a non-monotonic pattern over time, with oscillations between abstraction (reduction) and enrichment (expansion). Table 3c

TABLE 3
Degree of chunking/enrichment exhibited in control recipes

Interface	Subject	Recipe No. 1	2	3	4	5	6	Mean
(a) Length of control receipe (number of words)								
P	WL	65	131	176	115	189	87	127
	TL	115	184	205	152	197	149	167
	ML	47	56	69	72	66	57	61
P + F	IS	91	60	125	91	150	206	120
	AS	42	134	79	62	131	91	90
	AV	46	70	70	50	53	47	56
	Mean	68	106	121	90	131	106	
(b) Number of distinct steps in recipe								
P	WL	7	4	5	7	7	5	
	TL	7	5	3	—	5	5	
	ML	7	8	9	9	6	7	
P + F	IS	10	6	7	10	6	6	
	AS	6	9	6	7	7	6	
	AV	6	10	8	9	9	6	
(c) Average number of words per recipe step								
P	WL	9.3	32.8	35.2	16.4	27.0	17.4	
	TL	16.4	36.8	68.3	—	39.4	29.8	
	ML	6.7	7.0	7.7	8.0	11.0	8.1	
P + F	IS	9.1	10.0	17.9	9.1	25.0	34.3	
	AS	7.0	14.9	13.2	8.9	18.7	15.2	
	AV	7.7	7.0	8.8	5.6	5.9	7.8	

lists the average number of words per recipe step [Table 3(a) entries divided by Table 3(b) entries]. The non-monotonic changes over time found in previous analyses are observed here as well.

These non-monotonic changes in the length and grouping of the control recipes seem to be indicative of the dynamics of knowledge acquisition. During the experiment, subjects were apparently consolidating what they already knew, as well as gaining new knowledge and restructuring the way they controlled the system. Now that we have established that subjects were acquiring knowledge over the course of the experiment, we can try to investigate to what extent this knowledge was sensitive to particular goal-relevant features of the process.

4.4.2. Differentiation

One way to determine if subjects acquired useful, accurate knowledge is by determining the extent to which their control recipes became increasingly differentiated with respect to important properties of the process. A cognitive work analysis (CWA) of DURESS II (Vicente & Pawlak, 1994) was used to operationalize this question, since it identified a number of properties that subjects can learn about. Two of the most salient constraints that subjects can exploit are: (a) the fixed relationship between output demand and steady-state heater setting for each reservoir and (b) the redundant structure of the FWSs when neither demand exceeds 10 units/s. The control recipes were analysed for evidence of differentiation to each of these constraints.

The first constraint derives from the fact that the setpoint temperatures for the two reservoirs are constant. As a result, there is a simple pair of rules that one can use to determine what the heater setting for each reservoir should be set at to achieve the setpoint temperature. For Reservoir 1, the heater should be set at the same numerical value as the output demand for that reservoir (D1), whereas for Reservoir 2, the heater should be set at one-third of the numerical value of the output demand for that reservoir (D2). Table 4(a) shows the extent to which subjects verbalized either of these rules (note that subjects were not given these rules, and therefore, had to discover them for themselves). The results show that three subjects explicitly used these rules, two in the P group and one in the P + F group. The fact that these subjects differentiated between

TABLE 4

Degree of differentiation exhibited in control recipes

		Recipe no.					
Interface	Subject	1	2	3	4	5	6
(a) Number of references to steady-state heater setting as a ratio of reservoir demand, for each reservoir							
	WL	0	0	0	0	0	0
P	TL	0	0	2	2	2	2
	ML	0	0	0	2	2	2
	IS	0	0	0	0	2	2
P + F	AS	0	0	0	0	0	0
	AV	0	0	0	0	0	0
(b) Explicit references to decoupled FWS strategy							
	WL	0	0	0	0	0	0
P	TL	0	1*	1*	0	1*	1*
	ML	0	0	0	0	0	0
	IS	0	0	1*	0	1*	0
P + F	AS	0	0	0	0	0	0
	AV	0	1	1	1	1	1
(c) Number of asymmetrical statements							
	WL	0	0	0	1	0	0
P	TL	0	1	3	4	5	4
	ML	0	0	0	1	1	1
	IS	1	0	0	1	2	3
P + F	AS	0	0	0	0	0	0
	AV	0	0	0	0	1	2

Note: Asterisk (*) indicates subjects who mentioned pre-conditions for decoupling.

the two heaters in such a clear manner is a sign of knowledge acquisition. The results also show that TL discovered these two rules relatively quickly. This is consistent with the previous observation that TL was the most proficient and motivated of all the P subjects.

The second constraint that can be used to investigate differentiation is the strategy subjects used to configure the two FWSs. More specifically, a CWA of DURESS II (Vicente & Pawlak, 1994) revealed that when neither of the output demands exceeds 10 units/s, it is possible to simplify the control of water flow by decoupling the two FWSs. By keeping valves VA2 and VB1 shut, FWS A can be used to feed only Reservoir 1, and FWS B can be used to feed only Reservoir 2 (see Figure 1). This creates a one-to-one mapping between FWSs and reservoirs, thereby eliminating the crossover interactions that can exist if a many-to-many mapping strategy is used (i.e. all four valves open, with each FWS supplying both reservoirs). Table 4b shows how often subjects explicitly mentioned this decoupling strategy and the conditions under which it could be used. Three subjects eventually reported exploiting this strategy, one in the P group and two in the P + F group. Note that the subjects who were identified as being the most proficient in each group, TL and AV, reported using this strategy very early in the experiment. This differentiation between classes of demands is adaptive in that it allows subjects to adopt a simpler control strategy whenever possible.

Table 4(b) also shows how often each of these three subjects mentioned the preconditions for decoupling, or the conditions under which an alternate strategy is required. TL and IS did so consistently, specifying an alternative FWS configuration procedure that should be used if the precondition was not satisfied. In virtually every case, they expressed this alternative procedure in conditional form (e.g. IF [max D] > 10 THEN do alternative procedure). In contrast, AV never expressed the preconditions for the decoupling strategy. Instead, he wrote that the crossover valves (VA2 and VB1) should be used "if necessary". The significance of this difference will be addressed later.

Because DURESS II has a highly symmetrical structure, there is another natural way to investigate the effects of differentiation. For each pair of system components or subsystems that are symmetrically related (e.g. heaters, reservoirs, FWSs), one can observe whether subjects controlled them in the same way or differently. Because there are differences between symmetrical components or subsystems, it is adaptive to differentiate between them. Such actions are referred to as asymmetrical–like components or sub-systems are treated differently, in terms of control actions. Examples of asymmetrical actions include: having the initial setting for HTR1 higher than HTR2 because the setpoint for $T1$ is higher than that for $T2$; noticing that when a demand is greater than 10 it is always on Reservoir 2, and therefore that VA2 must be used in such situations, but that there is never a need to do the same with VB1; acting to stabilize $T2$ before $T1$ because the former will usually reach the setpoint (20°C) before the latter (40°C). Table 4(c) shows the frequency of such asymmetrical statements. The most salient result is that differentiation increased with practice, as one would expect. In the P group, we again see that TL exhibits, by far, the most differentiated control recipes. In the P + F group, the two best subjects, AV and IS, also exhibit asymmetries that are symptomatic of differentiation. The fact that the better subject, AV, exhibits less statements of this type compared to IS will be addressed in subsequent analyses.

In summary, these results indicate that, with practice, subjects increasingly differentiated between system components or subsystems in terms of how they acted on them. This differentiation is functional since it reflects, in virtually all cases, an adaptation to goal-relevant system constraints. In the P group, the most proficient subject, TL, consistently demonstrated a superior level of differentiation. In the P + F group, IS and AV showed more differentiation than the weaker subject, AS. Therefore, the subjects that were shown earlier to be the most competent also exhibited the greatest level of differentiation. However, knowledge acquisition and differentiation should not be equated with deep knowledge. Subjects may or may not understand the deep justification for their adaptive actions.

4.4.4. Knowledge organization

Knowledge acquisition can take many different forms, some deeper than others. For example, several qualitatively different types of knowledge organization can be postulated for the control recipes, including the following.

1. Knowledge tied to the perceptual features of the display.
2. Knowledge represented in the form of specific rote actions.
3. Knowledge indicating understanding of the rationale behind the recipes.
4. Knowledge that is tied to the system goals or functions that need to be achieved.

In this subsection, the control recipes will be analysed from each of these perspectives.

The first analysis examines the number of explicit references to perceptual features of the display. For example, subjects can refer to the green goal area, the vertical slope, the red line, etc. Table 5(a) shows the results of this analysis. The most salient finding is that the P + F group exhibited more statements of this type than the P group. This is not surprising since the P + F interface contains many more goal-relevant perceptual features that can be used for controlling the system (compare Figure 2 with Figure 1). Of the three P + F subjects, the best, AV, exhibits no references to perceptual features, whereas AS, the least competent, exhibits the most. IS exhibited an intermediate number of such references. The score for AS on the last recipe deserves a special comment. This recipe was written immediately after the transfer trials, and thus subjects were writing how they would control the system with the transfer interface (in the case of AS, the P interface). AS, however, spontaneously wrote two different recipes, one for each interface. Only the data for the transfer interface are included in Table 5(a), to make the results comparable to those of the other subjects. However, it is important to note that the recipe AS wrote for the P + F interface had many references to perceptual features of the display.

These results reinforce earlier findings indicating that the knowledge of AS is much more firmly entrenched in the surface features of the display than that of any other subject. This, alone, is not a bad thing since adaptation to perceptual features can also be accompanied by deep understanding. However, the fact that AS was the least competent of the P + F subjects and exhibited a very low degree of differentiation suggests that his knowledge was indeed relatively shallow. The remaining analyses in this subsection support this interpretation.

A second way in which subjects' control recipes can be organized is according to rote actions. This form of knowledge organization was evaluated by counting the number of statements that refer to specific, quantitative component settings. Note that references such as max, min, off, on, full, increase, decrease were not included. Only explicit quantitative references were counted (e.g. set the heaters to 50%; set VA1 to 10). Also, results for the first recipe were not included, since subjects did not yet have any

TABLE 5
Different types of knowledge organization exhibited in control recipes

		Recipe no.					
Interface	Subject	1	2	3	4	5	6
(a) Number of explicit references to perceptual features of the display							
P	WL	0	0	0	0	0	0
	TL	0	0	0	0	0	0
	ML	1	0	0	0	0	0
P + F	IS	0	0	0	0	2	0
	AS	0	2	2	3	1	0
	AV	0	0	0	0	0	0
(b) Number of statements specifying precise quantitative values for component settings							
P	WL	—	2	2	0	0	0
	TL	—	1	2	6	5	5
	ML	—	2	0	0	0	0
P + F	IS	—	1	1	1	1	1
	AS	—	0	0	0	0	0
	AV	—	0	0	0	0	0
(c) Number of declarative knowledge statements							
P	WL	0	0	1	1	0	0
	TL	0	0	0	0	0	0
	ML	0	0	0	0	0	0
P + F	IS	0	0	0	0	0	0
	AS	0	0	0	0	0	0
	AV	0	0	0	0	0	0
(d) Number of statements justifying or explaining recipe steps							
P	WL	0	2	1	1	3	0
	TL	0	3	2	0	0	0
	ML	0	0	0	0	0	0
P + F	IS	0	0	0	0	0	0
	AS	0	0	0	0	1	0
	AV	1	0	0	0	0	0
(e) Number of statements specifying the goal to be achieved, without listing the precise actions that need to be performed or the relationships that need to be considered							
P	WL	—	0	0	1	1	0
	TL	—	1	0	0	0	0
	ML	—	2	1	0	0	0
P + F	IS	—	1	1	1	0	0
	AS	—	0	0	0	0	2
	AV	—	1	1	2	2	2

experience to make meaningful references to quantitative control settings. Table 5(b) presents the results of this analysis. The most striking observation is that, once again, the best subject in the P group (TL) stands out, exhibiting a very large number of quantitative statements. In contrast, the best subject in the P + F group, AV, never specified any quantitative settings. These results suggest that TL's knowledge of the system is organized in a rote action-based manner, almost like a web of procedures; he seems to think predominantly at a very detailed level representing what he needs to do to the system.

A third frame of reference for organizing the control recipes is the degree to which subjects seem to understand the rationale behind their recipes. This dimension was analysed in two ways, first by examining the number of declarative knowledge statements made, and second, by examining the frequency with which subjects explained or justified their recipe steps. The results of the first analysis are presented in Table 5(c). Only one subject, WL, expressed (unsolicited) declarative knowledge statements in his control recipe, and even then, on only two occasions. This subject was in the P group. In the first reference, he describes the factors that affect the temperature response (volume, output flow rate, etc.), and in the second reference, he points out the fact that the relationship between valve settings and flow rates is usually not unique since various combinations of settings can produce the same set of flow rates.

Table 5(d) shows the results of an analysis examining how often subjects spontaneously justified the steps they had written, even though they were not asked to do so. For example, several subjects pointed out that they initially set the heaters to maximum, and then went on to explain that the reason for doing so was to make the temperature go up quickly. Several results emerge from this analysis. The P subjects, WL and TL, exhibited such statements more frequently than did the P + F subjects. Furthermore, WL provided explanations more consistently than TL. This reinforces the findings of the declarative knowledge analysis, suggesting that WL consistently tried to understand the basis for his control strategies. In contrast, TL initially provided a large number of explanations/justifications but then ceased to do so. Referring back to Tables 4(c) and 5(b), the point at which TL stops making explanations seems to correspond with the point at which he begins to increase the richness and density of his action-based statements. This leads to an intriguing hypothesis: at some point in the experiment, TL may have started to focus more on developing detailed procedures for controlling the system than on further developing his understanding of why the procedures are effective, or needed. Collectively, these results indicate that WL is the only subject who explicitly showed consistent signs of trying to reflect on and understand the basis for the procedures he was using (this does not rule out the possibility that other subjects did so as well, but simply did not write about it in their control recipes).

The final analysis of the control recipes examines a very different form of knowledge organization. Whereas the previous analyses have focused primarily on the actions that need to be taken, one can also organize one's knowledge around the system goals that need to be achieved. Table 5(e) analyses the control recipes from this perspective by counting the number of statements specifying only a goal to be achieved, without listing the specific actions that need to be performed to achieve the goal, or even the relationships that need to be considered in selecting the actions. Examples of this type of statements include: "set the valves to match the demand", "adjust the heater to stabilize temperature in goal state". Again, the first recipes are excluded from the analysis. Table 5(e) shows that the best of the P + F subjects, AV, thinks about the system in this manner. Starting at recipe 4, his responses stand out from those of the other subjects.

Rather than focusing exclusively on the specification of actions, a significant part of AV's control recipes are goal-based or function-based. This suggests that, for part of his control recipe, AV focuses on the goals to be achieved, and does not have a rigid set of actions that he uses every time to achieve those goals. Table 5(e) also shows that ML and especially IS exhibited several goal-based statements early on but then stopped doing so. In addition, WL generated several such statements near the end of the experiment.

The score of AS on recipe 6 deserves a special comment. Although he also specified his recipes in a function-based form, there is a subtle but important difference between the wording of AS and that used by AV. Whereas statements by AV are of the form "adjust (component) *to* (goal state) ", the statements by AS are of the form "adjust (component) *until* (goal-state) ". The former indicates a directed strategy whereas the latter suggests an iterative trial and error approach. This interpretation is confirmed, at a detailed level, by video tapes showing how these subjects control the system, and at a more global level, by the substantial difference in competency between them.

Table 5(b) describing how often subjects specified precise, quantitative values for component settings is also relevant to functional organization in a negative sense. That is, the specification of quantitative values for action is basically the antithesis of functional organization, where only the goals to be achieved are mentioned. Therefore, the more a subject exhibits one type of organization, the less they can exhibit the other, by definition. Table 5(b) shows that AV never exhibited any quantitative statements.

The analysis in Table 4(b) showing whether subjects adopted a decoupling strategy for configuring the FWSs is also pertinent to the identification of functional knowledge organization in a subtle yet interesting way. Recall that three subjects (TL, IS and AV) showed evidence of using this strategy. However, the first two, TL and IS, represented the preconditions for the strategy in an action-based form (IF [max D] > 10 THEN do alternate strategy). AV, on the other hand, stated that the crossover valves should be used "if necessary" to achieve the goal of interest. AV's statement is more consistent with a function-based knowledge organization since it does not specify under what conditions the crossover valves should be used, nor how they should be manipulated when required. Collectively, these results strongly suggest that AV's knowledge of the system is organized in a functional manner.

4.4.6. Discussion

Taken in isolation, none of the analyses just described can be given much weight. Each individual result can be interpreted in a myriad of ways. However, when considered as a whole and in the context of the performance data described earlier, the consistency of the findings begins to suggest a relatively coherent picture of the interaction between interface design and individual differences on subject's knowledge organization. This subsection will be devoted to elucidating the major themes in this picture. The discussion will be organized by subject, with a summary of each subject's responses. In addition, the last recipe written by each subject for the interface they used throughout the vast majority of the experiment has been transcribed (see Tables 6–11) to give a concrete example of the patterns being summarized (for 5 out of the 6 subjects, this recipe is number 5; for AS, the part of recipe number 6 that was devoted to the P + F interface is presented since it is more representative of his overall pattern of responses). These transcriptions are as literal as possible, retaining typographic features (e.g. indenting, line length) and any other idiosyncratic details (e.g. spelling and grammatical errors) contained in the original hand-written recipes. The responses of the subjects in the P interface will be discussed first.

Table 6 shows the last control recipe written by TL for the P interface. The most distinguishing features in TL's recipes are the converging signs pointing to a rote, action-based knowledge organization scheme. This is shown in Table 6 in the great level of detail and the precision with which actions were specified. With experience, TL was able to develop a rich set of procedures for operating the system. Judging by his performance they served him very well, at least for normal trials. Another prominent feature of TL's recipes is the strong degree of differentiation. TL quickly learned to pick up on system constraints and exploit them in his control recipes. This is consistent with the fact that TL was deemed to be the best subject in the P group.

However, adaptation should not be equated with understanding. TL exhibited no declarative knowledge statements in his recipes, and while he provided explanations early on, he stopped doing so at about the same time that he began to further develop very detailed actions for controlling the system. As mentioned earlier, this can be interpreted as trading off deep understanding for procedural efficiency. In fact, there is additional evidence to support this interpretation. Table 6 shows that, after pointing out the rules for obtaining the steady state heater settings as a function of demand, TL points out: "It might not make sense, but it works"! What is particularly striking, however, is that despite this focus on very detailed actions (perhaps to the detriment of understanding), TL was the most competent of the P subjects. Having said that, it is important to recognize that TL's performance on fault trials, especially non-routine faults, was quite poor compared to AV, showing the brittleness of TL's action-based adaptation.

TABLE 6

Last control recipe for TL on P interface

---

(1) Open valves VB2 and VB to their maximum (10) and start PB. Set HTR2 to 4.5.
(2) Set VA1 to 7, VA to maximum (10), and start PA. Set HTR1 to 10.
(3) As the temperature in Res 2 reaches the lower end of the desired range, set HTR2 to 3 1/3; then set VO2 to the desired level.
(4) Set VO1 to the desired level. As the temperature in Res 1 reaches the lower end of the desired temperature range, set VA1 to level 10.
(5) As the reservoirs reach level 60, adjust VA1 to equal VO1 and VB2 to equal VO2. Modify heater settings as necessary.

Hint: HTR1 should be at the same numerical setting as VA1, and HTR2 at 1/3 of VB2. (It might not make sense, but it works).

IF THE DEMAND ON RES 2 *IS GREATER THAN* 10
Follow steps (1)–(3) above.
(4) Set VA2 such that the total water input into Res 2 is one level above VO2. Set VA1 such that VA1 + VA2 = 10. Set heater levels as defined in the hint section above.
(5) As the reservoirs reach level 60, adjust values so that VA1 = VO1 and VA2 + VB2 = VO2. Modify heater settings as required.

TABLE 7
Last control recipe for WL on P interface

1. Click OK button. Duress simulator pops up ready for use.
2. Set VA, VB, VA1, VA2, VB1, VB2 to maximum, VO1 and VO2 to minimum.
3. Turn on pumps PA and PB so that water comes in.
4. As long as there is any water in the reservoirs RES1 and RES2, turn on the heater HTR1, HTR2 to maximum so that the temperature of the water in the reservoir could rise quickly to the desired value.
5. When the water level almost reaches half capacity of the reservoirs, adjust VA1, VA2 and VB1, VB2 to allow input water flow equal output flow. Immediately, adjust the heater to allow temperature of water in the reservoirs become stable there within the desired temperature range.
6. Usually, some subtle change of heater is necessary to keep the system in equilibrium state. With some experience dealing with heater setting, usually the user could quickly set the right level.
7. Some exceptions may occur, for example, change in input temperature, desired output rate is higher than allowed maximum input flow. Again the key point is still to focus on temperature control and water flow rate control.

Table 7 shows the last control recipe written by WL for the P interface. The most distinguishing feature of WL's recipes is the consistent emphasis on declarative knowledge statements, and especially statements justifying or explaining the rationale for various steps in the recipe. The latter feature can be observed in Table 7 in two places. WL was unique among all subjects in the emphasis his recipes showed on understanding. However, he exhibited virtually no signs of differentiation, and while he started off by making reference to specific quantitative settings, after a while he stopped altogether. In fact, towards the end of the experiment, WL shows some evidence of function-based thinking, the very antithesis of specifying precise quantitative settings.

It is instructive to point out similarities and differences between WL and TL. Both started off by providing several comments justifying the steps in their recipes and by making some reference to specific quantitative settings. However, the similarities end there. While TL went on to adopt a heavily action-based perspective, WL continued to exhibit more emphasis on understanding the rationale for his actions and even showed some evidence of function-based thinking. Thus, despite the initial similarities, these two subjects wound up going in very different directions (cf. Tables 6 and 7). One might expect that WL would perform better since he seemed to focus more on higher level understanding rather than low level actions, but in fact, the previous results show that TL was by far the more competent of the two subjects.

Table 8 shows the last control recipe written by ML for the P interface. Compared to most of the other subjects, ML's recipes do not show many pronounced characteristics. Interestingly, he started off by exhibiting some function-based steps but then ceased to do so. It was at this point that he started showing some signs of differentiation (although nowhere near as much as TL). The other significant feature about ML which is not captured by any of the preceding analyses, but which can be seen in Table 8, is the process that he used to configure the FWSs. His strategy includes all of the valves and

TABLE 8

Last control recipe for ML on P interface

(1) Open input valves
(2) Turn on pumps
(3) Turn on heaters
(4) Adjust input valve settings so that
    — input adds to output
    — input is proportioned according to output demands
(5) Adjust heaters to achieve demand temperature
    • upper heater should be set so that heater setting = output setting
    • lower heater should be set so that heater setting = 1/3 output setting
(6) Wait for steady state and adjust (minor) where necessary

requires him to do a fair bit of mental arithmetic (e.g. computation of ratios). This may be one of the reasons why he was the least competent of the P subjects.

Table 9 shows the last control recipe written by AS for the P + F interface. The recipes of AS were distinctive, not only among the P + F subjects, but overall as well. He was the only subject to consistently show evidence of relying on the perceptual features of the P + F interface. This pattern is very clear in Table 9. In addition, he showed virtually no evidence of differentiation or declarative or explanatory statements. These results suggest that the control strategies of AS were highly dependent on the surface features of the display and that he did not possess a great deal of deeper knowledge. The fact that he was the least competent of the P + F subjects is consistent with this interpretation. There is also some weak evidence in the recipes that AS may have been using a trial and error control strategy, symptomatic of a shallow knowledge base.

Table 10 shows the last control recipe written by AV for the P + F interface. Like other competent subjects, AV exhibited some signs of differentiation. However, his distinguishing feature was his emphasis on function-based statements in the latter part of his recipes. This result, and others showing a comparative absence of action-based steps, suggest that AV focused more on the goals that he needed to achieve rather than on exactly how he achieved them. The only exceptions are the relatively well defined initial steps of his recipes (see Table 10), which are similar to those of other subjects. Thus, AV can be viewed as the antithesis of TL. In fact, the contrast between TL's recipe in Table 6 and AV's in Table 10 could not be stronger. This is surprising since these two subjects were the best performers in each of their groups. The fact that the knowledge organization of the best P + F subject was so markedly different from that of the best P subject suggests that the interface subjects are using has a significant influence in shaping subjects' behaviour and way of thinking about the process.

Table 11 shows the last control recipe written by IS for the P + F interface. Like AV, IS started off by exhibiting evidence of function-based thinking. However, these signs were replaced over time by an increasingly strong evidence of action-based steps and some evidence of perceptual-based adaptation. IS also exhibited signs of differentiation, particularly towards the end of the experiment. Like many other subjects, IS did not show any explicit signs of justifying the steps in his recipes.

Table 9

Last control recipe for AS on P + F interface

(1) OPEN VALVES VA, VB, VA1, VA2, VB1 AND VB2 TO MAXIMUM
(2) SWITCH ON BOTH PUMPS PA AND PB
(3) ONCE THERE IS WATER IN THE RESERVOIRS, SWITCH ON THE HEATERS HTR1 AND HTR2.
(4) WHEN THE RESERVOIRS ARE A QUARTER FULL, TURN ON VO1 AND VO2 AND SET THEM TO MATCH THE SETPOINTS
(5) ADJUST VA AND VB SUCH THAT VA + VB = VO1 + VO2
(6) ADJUST VA1 TO VB2 UNTIL FLOW RATES IN BOTH RES1 AND RES2 REACH STEADY STATE (THE GREEN LINES ARE VERTICAL)
(7) ADJUST HTR1 AND HTR2 UNTIL THE RED LINES ARE VERTICAL IN BOTH RESERVOIRS AND THE CENTRE RED LINES ARE WITHIN THE GREEN BARS (SETPOINTS)
(8) AT THIS POINT, THE SYSTEM HAS REACHED STEADY STATE.

Table 10

Last control recipe for AV on P + F interface (experimenter notes are in brace brackets)

Turn PA & PB ON
Turn VA & VA1 ON (MAX)
Turn HTR1 ON (MAX)
Turn VB & VB1 & VA2 (if necessary) {arrow underneath text, from parenthetical phrase to VA2}
Turn HTR2 ON (MAX)
Adjust VO1 & VO2 TO OUTPUT
Adjust VA & VA1, VB & VB1 & VA2 to necessary input.
Adjust HTR1 & HTR2 For the INPUTS
DO SOME FINE TOUNING. END

Because of the lack of distinctive patterns, it is difficult to characterize the knowledge organization of IS. In some ways, he was similar to TL in that he showed an increasing tendency towards action-based organization and differentiation, although not nearly as strongly as TL did. He was also weakly similar to AS in his references to perceptual features in recipe 5. Perhaps the surest conclusion is that IS was clearly different from AV, since he showed no evidence of function-based organization by the end of the experiment (cf. Tables 10 and 11).

5. General discussion

What have we learned about how interface design and individual differences in learning motivation interact to affect deep knowledge? Beginning with the P interface, it seems that to do well, subjects had to adopt a rote, proceduralized knowledge structure. TL, who was the most proficient subject in the P group, exhibited very clear signs in his

TABLE 11

Last control recipe for IS on P + F interface

— Turn on valves VA, VA1, VB, VB2 to maximum (10)
— Turn on Pumps PA & PB
— Turn on Heaters; Heater 1 to maximum, Heater 2 to 1/2
— Set output valves to proper values (indicated by green bar). The setting for VO1 is always < 10. If the setting for VO2 is > 10, allow additional H_2O to enter the resiviour via valve VA2 to make up the difference.
— Set input valves to match output valves
  VA1 + VB1 = VO1
  VB2 + VA2 = VO2
  VO1 + VO2 *must be* ≤ 20
— Set heater setting so that temp stabilizes (energy in = energy out), and red line is vertical. As a rough guide, for resiviour 1, the temp setting should = output setting. Example: output = 8, then heater should be ~8. For resiviour 2, heater setting should be 1/2 output setting (or less)
  Output = 8
  heater ≤ 4

control recipes of having a highly differentiated set of rules that allowed him to control the system very efficiently during normal trials. However, as he himself implied, this is not equivalent to understanding why those rules work. The only subject in the group who seemed to try to understand these relationships throughout the experiment was WL, and he was not nearly as proficient as TL. Interestingly, it seems that TL, who was highly motivated to do well, started off by trying to understand the basis for his control strategies, but eventually gave up that strategy in favour of developing a highly proceduralized strategy. With such a small sample size, it is difficult to know for certain, but the P interface may make it very difficult for subjects to perceive, comprehend, and consistently take into account the underlying relationships governing the behavior of the process. As a result, adopting a rote proceduralized knowledge organization may be the most psychologically feasible way to succeed with the P interface. The fact that fault management performance was generally worse with the P interface, even for TL, shows that a dependency on rules is a relatively shallow form of knowledge, no matter how efficient it is for normal trials.

The effects of the P + F interface are very interesting since they vary quite considerably from subject to subject, but in a coherent way. AS was consistently the least competent of the P + F subjects throughout the experiment. It was noted that he did not reflect on the task, and was not highly motivated to do well. It is not surprising then that AS explicitly mentioned that he controlled the system through reliance on the graphics, without thinking about what the graphics represented. This was very clear in his control recipes, where he repeatedly referred to the perceptual features of the display. This surface approach to learning is consistent with the performance criteria that seemed to drive AS; by doing the task this way, he did not have to think very much. However, this had enormous implications for his performance. Not only was he the least proficient of the P + F subjects, he actually got worse on some performance measures as the experiment evolved (Christoffersen *et al*., 1996). And when he transferred to the P interface and had to control the system without the familiar graphics, his performance completely broke down after one trial—he simply could not do the task, even in the absence of a fault. Thus, it seems that the P + F interface, when combined with a reluctance to think about and reflect upon the feedback that the interface provides, can lead to a dependency on the perceptual features of the display. This dependency, in turn, leads to poor performance and a very shallow knowledge base. It is essential to note, however, that the performance of AS on normal and fault trials as a whole was no worse—and if anything, slightly better—than that of ML, who was also not well motivated. Thus, given a lack of motivation, the P interface does not provide a very attractive alternative to the P + F interface.

AV, who was the most proficient subject in the P + F group on most measures and tasks, adopted a way of thinking about the process that was essentially the antithesis of knowledge organization centred around specific rote actions. Although his initial actions during startup followed a set pattern (like all other subjects), after this phase, his control recipes indicated that he thought about the process in terms of goals or functions to be achieved. This is precisely the kind of thinking that the P + F interface is intended to engender (Vicente & Rasmussen, 1992). Because the interface provides a visualization of the process constraints, subjects do not need to rely strictly on procedures. Rather, they can use the feedback provided by the interface as an error signal for controlling the process. From this viewpoint, what matters is that the goals are achieved, not how. A familiar example is keeping one's car between lane markers while driving. Because the state of the system (i.e. the position of the car) is clearly visible with respect to the goal (i.e. the lane markers), experienced drivers do not have to memorize a procedure for how to control the car. Instead, they can just rely on the feedback from the environment to guide their actions directly in a goal-directed fashion. In fact, one could argue that it would be counterproductive to memorize a procedure because it would not allow one to navigate effectively in the presence of unanticipated events (e.g. wind gusts), whereas a function-based approach should, since the goal is always to nullify the error. The only thing that changes from situation to situation is how this goal should be accomplished. Note that the function-based knowledge organization is very difficult to adopt in the absence of rich feedback, since it would take a great deal of computation on the part of the P subjects to generate the information that is being provided directly by the P + F interface (Christoffersen *et al*., 1997).

6. Conclusions

Wickens (1992) has argued that interfaces that are based on graphical representations should lead to an improvement in performance. Previous analyses of performance on normal and fault trials, briefly summarized here, are consistent with this observation. An interface based on EID can lead to more consistent performance on normal trials as well as more accurate and more efficient fault management performance compared to a more traditional interface format (Christoffersen *et al*., 1996, 1997). Wickens has also argued, however, that this performance benefit may come at the cost of a deficit in long-term retention of knowledge. Presenting information in an externalized graphical representation, as EID tries to do, may allow subjects to perform domains tasks without engaging in deep processing, thereby leading to a shallow knowledge base.

Before discussing how our findings bear on this issue, it is worthwhile reviewing the general conditions under which our experiment was conducted. First, the study was carried out over six months to ensure that subjects would have a chance to become adapted to the perceptual features of each interface. Second, the knowledge elicitation tests were administered infrequently, so as not to cause subjects to reflect on their performance more than they otherwise would. Third, no training was given, so that subjects had to engage in discovery learning. Fourth, non-routine faults were introduced, but only at the end of the experiment. All of these factors would seem to improve the chances that the phenomenon identified by Wickens would appear. After all, in an industrial setting, there are many forces which would encourage operators to reflect upon, and process, information more deeply than in this experiment (e.g. they may have to communicate their decisions to others, they are accountable for their actions, training provides them with a basis for control that goes beyond mere perceptual features of the interface, the potential safety hazard is obviously much greater and their livelihood depends on their performance). Thus, there are several reasons to suggest that this experiment was conducted under conservative conditions.

A weak version of Wickens' argument would merely suggest that some P + F subjects could become dependent on the perceptual features of the display, and that this would negatively affect their knowledge base. There was clear support for this hypothesis. AS did in fact use the graphics in the P + F interface as a crutch. Not only did he exhibit a shallow knowledge base, but his performance was also markedly worse than the other two subjects in his group. Furthermore, it was not unusual for the best subject in the P group, TL, to outperform AS, although this occurred more frequently during normal trials than during faults. These results show that, given a lack of motivation and/or a surface approach to learning, an EID interface can lead to a shallow knowledge base that translates into comparatively weak performance.

One might ask, what is the alternative? At this point, it is useful to consider the strong version of Wickens' (1992) argument, which would suggest that most or all P + F subjects would actually have a shallower knowledge base than the P subjects, for the reasons mentioned earlier. There is no evidence to support this hypothesis. On the contrary, even the best of the P subjects exhibited a shallow knowledge organization based on a rote set of procedures which were not accompanied by a deep understanding of the process. In contrast, the best of the P + F subjects exhibited a functionally organized knowledge base that is consistent with the aims of the EID framework. These differences in knowledge structures were accompanied by corresponding differences in fault management performance. Most of the P + F subjects performed better than the best P subject during routine and non-routine faults (Christoffersen *et al.*, 1997). In no trial did the best P subject outperform the best P + F subject in terms of fault management performance. And if one matches subjects in the two groups according to perceived level of motivation and attitude towards learning, one finds that the subject in the P + F group generally outperformed his cohort. Therefore, while it is possible for subjects to have a shallow knowledge base and perform poorly with the P + F interface, it seems that a subject with similar characteristics would, if anything, do worse with the

P interface. Based on these findings, there seems little reason to choose traditional interface design practices over EID. Thus, the strong version of Wickens' (1992) argument does not find any support. A trade-off between performance and deep knowledge does not seem to exist.

This conclusion is of obvious theoretical interest, but it has important practical implications as well. If generalizable, it means that designing an interface to improve performance can result in improvements in deep knowledge as well. This is very good news for embedded training applications, since it suggests that the same interface can be used by operators to perform domain tasks and to refresh their domain knowledge as well.

6.1. LIMITATIONS

There are several limitations to this study which must be taken into account in generalizing the findings obtained here to other contexts. First, although the experiment was deliberately designed to be representative of the situations to which we are interested in generalizing (the control of complex industrial systems), we cannot be certain that the results we obtain will generalize readily to industry-scale systems. Besides the obvious differences in scale, in this experiment, subjects did not have procedures or alarms to guide them in dealing with faults. Furthermore, they were not provided with any training, and had to learn about the system in a discovery-based fashion through extensive practice at controlling the system. It is not known to what extent these differences could affect the generalizability of the results obtained here. Thus, it is important to replicate this study using a more complex simulation facility (e.g. Yamaguchi, Furukawa & Tanabe, 1997).

Second, because of the great emphasis on representativeness, we had to trade-off experimental control and statistical power. Thus, we do not have any evidence of the reliability of our findings. We have tried to compensate for this by adopting multiple, converging measures and by analysing each subjects' performance individually and in detail. To appreciate the contribution of this work, it is essential that it be evaluated by the criteria that it is intended to satisfy, not those that are typically used to evaluate traditional laboratory studies.

Third, our frank description of the subjects' motivations might give some readers cause for concern. It is very rare that one finds an experiment where a subject reads a magazine while he is performing the required tasks, where subjects' actions can damage the system (and thereby end the trial prematurely), where subjects sacrifice performance in order to deliberately explore the safety boundaries of the system, and where some subjects give up on their task because it is too effortful or frustrating. However, all of these things can, and do, occur in actual work situations. Thus, whereas our experiment may lack experimental control, it is certainly far more representative of operational settings than the usual laboratory experiment. Thus, it provides valuable insights that complement those of traditional experiments (cf. Vicente, 1997).

6.2. FUTURE RESEARCH

The strong set of individual differences observed in each interface group was one of the most salient results from this experiment. In particular, not all of the P + F subjects were able to take advantage of the benefits of EID. Are there cognitive styles or abilities that inherently limit an individual's performance with an interface designed according to the principles of EID? Is there something that can be done to bring up the performance of subjects, such as AS, who do not do as well if left on their own, as in this experiment? These questions are not only of theoretical interest, but they are of great practical interest as well, since they have very strong implications for training or selection. Future research should determine which of these avenues is the most promising for ensuring that operators can make the most of interfaces based on EID.

This research was sponsored by a contract from the Japan Atomic Energy Research Institute (Dr Fumiya Tanabe, Contract Monitor), as well as a graduate scholarship and research and equipment grants from the Natural Sciences and Engineering Research Council of Canada. We would like to thank Dimitrios Nastos, Andy Sun, Peter Benda, Dianne Howie, and Dr Tanabe for their contributions to this research, and the anonymous reviewers for very helpful suggestions. Finally, we are extremely grateful to the six subjects who consistently devoted so much of their time to this experiment. Without their cooperation, this study would not have been possible.

References

BECKER, G. (1991). Analysis of human behaviour during NPP incidents: a case study. In *Balancing Automation and Human Action in Nuclear Power Plants*, pp. 517–526. Vienna, Austria: IAEA.

BENNETT, K. B. & FLACH, J. M. (1992). Graphical displays: implications for divided attention, focused attention, and problem solving. *Human Factors*, **34,** 513–533.

BISANTZ, A. M. & VICENTE, K. J. (1994). Making the abstraction hierarchy concrete. *International Journal of Human–Computer Studies*. **40,** 83–117.

BRUNSWIK, E. (1956). *Perception and the Representative Design of Psychological Experiments*. 2nd edn. Berkeley, CA, USA: University of California Press.

CHRISTOFFERSEN, K., HUNTER, C. N. & VICENTE, K. J. (1994). *Research on factors influencing human cognitive behaviour* (I) CEL 94-05, Cognitive Engineering Laboratory, University of Toronto, Toronto.

CHRISTOFFERSEN, K., HUNTER, C. N. & VICENTE, K. J. (1996). A longitudinal study of the effects of ecological interface design on skill acquisition. *Human Factors* **38,** 523–541.

CHRISTOFFERSEN, K., HUNTER, C. N. & VICENTE, K. J. (1997). A longitudinal study of the effects of ecological interface design on fault management performance. *International Journal of Cognitive Ergonomics* **1,** 1–24.

COEKIN, J. A. (1969). A versatile presentation of parameters for rapid recognition of total state. In *Proceedings of the IEE International Symposium on Man–Machine Systems*. Cambridge, England: IEE.

GOODSTEIN, L. P. (1981). Discriminative display support for process operators. In J. RASMUSSEN & W. B. ROUSE, Eds. *Human Detection and Diagnosis of System Failures*, pp. 433– 449. New York, USA, Plenum.

HOWIE, D. E. (1996). *Shaping expertise through ecological interface design: strategies, metacognition, and individual differences*, Unpublished Master's Thesis, Department of Education, University of Toronto, Toronto.

IRMER, C. & REASON, J. T. (1991). Early learning in simulated forest fire-fighting. In *Simulations, Evaluations, and Models: Proceedings of the 4th MOHAWC Workshop*, pp. 1–20. Risø National Laboratory, Roskilde, Denmark.

JANZEN, M. E. & VICENTE, K. J. (in press). Attention allocation within the abstraction hierarchy. *International Journal of Human–Computer Studies*.

KIERAS, D. E. & BOVAIR, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science*, **8,** 255–273.

Kossack, M. F. (1992). *Ecological task analysis: a method for display enhancement*. Unpublished Master's Thesis, Georgia Institute of Technology, Atlanta, GA, USA.

Maddox, M. E. (1996). Critique of "A longitudinal study of the effects of ecological interface design on skill acquisition" by Christoffersen, Hunter, and Vicente. *Human Factors*, **38,** 542–545.

Malone, T. B., Kirkpatrick, M., Mallory, K., Eike, D., Johnson, J. H. & Walker, R. W. (1980). *Human factors evaluation of control room design and operator performance at Three Mile Island*—2, NUREG/CR-1270, USNRC, Washington, DC.

Moore, P. J. Smith, M. W. & Telfer, R. A. (1994). Learning for automation: generic versus specific approaches. In M. Mouloua & R. Parasuraman eds. *Human Performance in Automated Systems*: *Current Research and Trends*, pp. 307–313. Hillsdale, NJ, USA: Erlbaum.

Mumaw, R. J., Woods, D. D. & Eastman, M. C. (1992). *Interim report on techniques and principles for computer-based display of data*, STC Report 92-ISJ4-CHICR-R1, Westinghouse Science and Technology Center, Pittsburgh, PA.

Pawlak, W. S. & Vicente, K. J. (1996). Including effective operator control through ecological interface design. *International Journal of Human–Computer Studies*. **44,** 653–688.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon, Ed. *Distributed Cognitions: Psychological and Educational Considerations*, pp. 47–87. Cambridge, UK: Cambridge University Press.

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decisionmaking and system management. *IEEE Transactions on Systems. Man. and Cybernetics*. **SMC-15,** 234–243.

Salomon, G. (1993). No distribution without individuals' cognition: a dynamic interactional view. In G. Salomon, Ed. *Distributed Cognitions*: *Psychological and Educational Considerations*, pp. 111–138. Cambridge, UK: Cambridge University Press.

Seminara, J. L., Gonzalez, W. R. & Parsons, S. O. (1997). *Human factors review of nuclear power plant control room design*, EPRI NP-309. Electric Power Research Institute, Palo Alto, CA.

Sharp, T. D. (1996). *Progress towards a development methodology for decision support system for use in time-critical, highly uncertain, and complex environments*. Unpublished Doctoral Dissertation, Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati, OH, USA.

Vicente, K. J. (1991). *Supporting knowledge-based behaviour through ecological interface design.* Unpublished Doctoral Dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

Vicente, K. J. (1997). Heeding the legacy of Meister, Brunswik, and Gibson: toward a broader view of human factors research. *Human Factors*, **39,** 323–328.

Vicente, K. J., Burns, C. M., Mumaw, R. J. & Roth, E. M. (1996). How do operators monitor a nuclear power plant? A field study. In *Proceedings of the 1996 American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human–Machine Interface Technologies*, pp. 1127–1134. LaGrange Park, IL, USA: ANS.

Vicente, K. J., Christoffersen, K. & Hunter, C. N. (1996). Response to Maddox critique. *Human Factors*, **38,** 546–549.

Vicente, K. J., Christoffersen, K. & Pereklita, A. (1995). Supporting operator problem solving through ecological interface design. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-25,** 529–545.

Vicente, K. J. & Pawlak, W. S. (1994). *Cognitive work analysis of the DURESS II system.* Technical Report CEL 94-03, Cognitive Engineering Laboratory, University of Toronto, Toronto, Canada.

Vicente, K. J. & Rasmussen, J. (1990). The ecology of human–machine systems II: mediating "direct perception" in complex work domains. *Ecological Psychology*, **2,** 207–249.

Vicente, K. J. & Rasmussen, J. (1992). Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-22,** 589–606.

Wickens, C. D. (1992). Virtual reality and education. In *Proceedings of the 1992 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 842–847. Piscataway, NJ, USA: IEEE.

WOODS, D. D. (1991). The cognitive engineering of problem representations. In J. ALTY & G. WEIR, Eds. *Human–Computer Interaction and Complex Systems*, pp. 169–188. London: Academic Press.

XU, W. (1996). *Externalizing a work domain structure on a hypertext interface using an abstraction hierarchy: Supporting complex search tasks and problem solving activities.* Unpublised Doctoral Dissertation, Department of Psychology, Miami University, Oxford, OH.

YAMAGUCHI, Y., FURUKAWA, H. & TANABE, F. (1997). JAERI's activities on simulator experiment evaluation methodologies and its facility. In *Proceedings of the IEEE 6th Conference on Human Factors and Power Plants*. Piscataway, NJ, USA: IEEE.

Paper accepted for publication by Associate Editor, Dr. R. Agarwal.